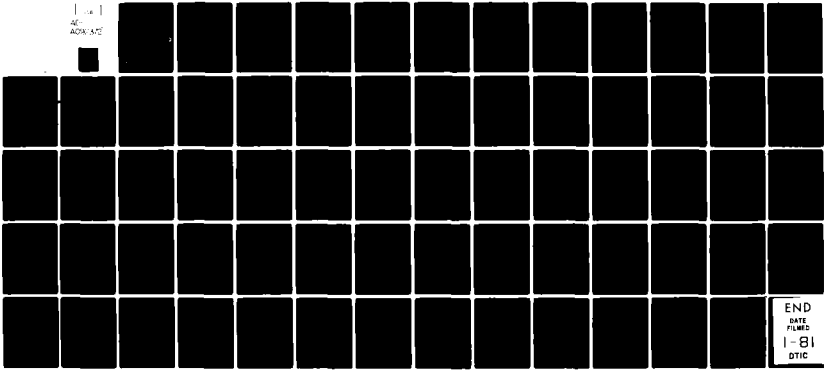
AD-A092 372 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH F/G 4/2
DETERMINATION OF THE SYSTEMATIC BIAS AND VARIABILITY OF THE ERR--ETC(U)
1980 J K WOESSNER
UNCLASSIFIED AFIT-CI-80-40T NL



UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 80-40T	2. GOVT ACCESSION NO.: AD-A092372
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Determination of the Systematic Bias and Variability of the Error of the National Meteorological Center's Six-Layer Baroclinic Model	5. TYPE OF REPORT & PERIOD COVERED: THESIS/DISSERTATION
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): James Keil Woessner	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT: University of Washington	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR, WPAFB OH 45433	12. REPORT DATE: 1980
	13. NUMBER OF PAGES: 55
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASS
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC ELECTE DEC 3 1980 S D A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-17

FREDRIC C. LYNCH, Major, USAF
Director of Public Affairs

Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

80 11 24 158

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A092372

DDC FILE COPY

Abstract

Determination of the Systematic Bias and
Variability of the Error of the
National Meteorological Center's
Six-Layer Baroclinic Model

by James Keil Woessner
Captain, USAF

Awarded
Master of Science
from
University of Washington, 1980

The geographical distribution of the bias and variability of the error in the 500 and 1000mb geopotential height forecasts produced by the National Mereorological Center's operational six-layer baroclinic model was investigated. Forecasts and the verifying observations for the six winter seasons beginning with the 1973-74 winter were used. It was found that the model systematically damps out the amplitude of the climatological mean surface highs and lows, and 500mb troughs and ridges. Typical values of the bias in the mean 500mb standing wave troughs are on the order of 60 meters; in these regions the bias accounts for up to 30% of the ~~root~~ mean square error. Throughout the forecast period, the standard deviation of the forecast error about its own mean shows regional contrasts similiar to ~~the~~ climatological standard deviation patterns. These non-systematic forecast errors tend to be largest in ~~the~~ "storm tracks" over the western oceans and quite small at low latitudes. Normalization of the standard deviation of the error by the climatological standard deviation of the observation results in a relatively featureless field except for a monotonic pole-to-equator increase. The normalized errors are larger at the 1000mb level than at the 500mb level over most of the hemisphere but they are of comparable magnitude over the eastern Alantic and Europe.

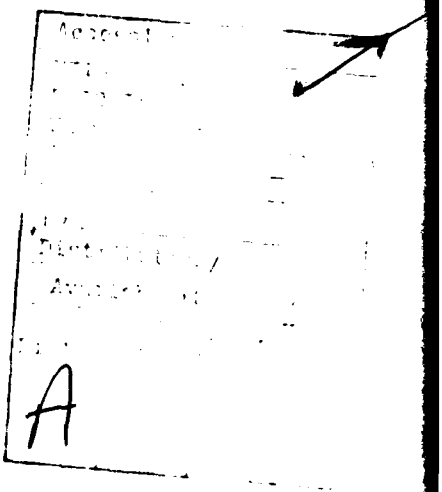
The vertical structure of the error field, as manifested in the geographical distributions of the correlation between the 1000 and 500mb forecast error, and the ratio of the amplitudes of error fluctuations at those levels, was found to be similiar to that of real atmospheric fluctuations.

An investigation of phase errors based on cross spectral analysis between time series of 72-hour forecasts and concurrent observations showed a small but systematic tendency for the model to underestimate the phase propagation of weather disturbances. The effect was not considered to be large or systematic enough to warrant empirical corrections.

DETERMINATION OF THE SYSTEMATIC BIAS AND VARIABILITY OF THE ERROR OF THE NATIONAL METEOROLOGICAL CENTER'S SIX-LAYER BAROCLINIC MODEL

by

JAMES KEIL WOESSNER



A thesis submitted in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE

UNIVERSITY OF WASHINGTON

1980

Approved by _____
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree _____

Date _____

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to ascertain the value and/or contribution of research accomplished by students or faculty of the Air Force Institute of Technology (ATC). It would be greatly appreciated if you would complete the following questionnaire and return it to:

AFIT/NR
Wright-Patterson AFB OH 45433

Research Title: Determination of the Systematic Bias and Variability of the Error of the National Meteorological Center's Six-Layer Baroclinic Model

Author: James Keil Woessner

Research Assessment Questions:

1. Did this research contribute to a current Air Force project?

   a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not?

   a. Yes b. No

3. The benefits of AFIT research can often be expressed by the equivalent value that your agency achieved/received by virtue of AFIT performing the research. Can you estimate what this research would have cost if it had been accomplished under contract or if it had been done in-house in terms of manpower and/or dollars?

   a. Man-years _____. b. $_____

4. Often it is not possible to attach equivalent dollar values to research, although the results of the research may, in fact, be important. Whether or not you were able to establish an equivalent value for this research (3 above), what is your estimate of its significance?

   2. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. AFIT welcomes any further comments you may have on the above questions, or any additional details concerning the current application, future potential, or other value of this research. Please use the back of this questionnaire for your statement(s).

NAME GRADE POSITION

ORGANIZATION LOCATION

USAF SCN 75-20B

University of Washington

Abstract

DETERMINATION OF THE SYSTEMATIC BIAS AND VARIABILITY OF THE ERROR OF THE NATIONAL METEOROLOGICAL CENTER'S SIX-LAYER BAROCLINIC MODEL

By James Keil Woessner

Chairperson of the Supervisory Committee:

Professor John M. Wallace
Department of Atmospheric Sciences

The geographical distribution of the bias and variability of the error in the 500 and 1000 mb geopotential height forecasts produced by the National Meteorological Center's operational six-layer baroclinic model was investigated. Forecasts and the verifying observations for the six winter seasons beginning with the 1973-74 winter were used. It was found that the model systematically damps out the amplitude of the climatological mean surface highs and lows, and 500 mb troughs and ridges. Typical values of the bias in the mean 500 mb standing wave troughs are on the order of 60 meters; in these regions the bias accounts for up to 30% of the root mean square error. Throughout the forecast period, the standard deviation of the forecast error about its own mean shows regional contrasts similar to the climatological standard deviation patterns. These non-systematic forecast errors tend to be largest in the "storm tracks" over the western oceans and quite small at low latitudes. Normalization of the standard deviation of the error by the climatological standard deviation of the observation

results in a relatively featureless field except for a monotonic pole-to-equator increase. The normalized errors are larger at the 1000 mb level than at the 500 mb level over most of the hemisphere but they are of comparable magnitude over the eastern Atlantic and Europe.

The vertical structure of the error field, as manifested in the geographical distributions of the correlation between the 1000 and 500 mb forecast error, and the ratio of the amplitudes of error fluctuations at those levels, was found to be similar to that of real atmospheric fluctuations.

An investigation of phase errors based on cross spectral analysis between time series of 72-hour forecasts and concurrent observations showed a small but systematic tendency for the model to underestimate the phase propagation of weather disturbances. The effect was not considered to be large or systematic enough to warrant empirical corrections.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I wish to thank John M. Wallace for suggesting this topic and for his guidance and critical insight throughout the course of this project. He has made my experience at the University of Washington both challenging and personally rewarding. I wish to express my appreciation to Harold J. Edmon and Ernest E. Recker for their assistance with computer graphics and programming. Thanks are also due to Lisa Link who typed the manuscript in a professional manner.

The data set used in the study was obtained from NCAR with the assistance of Dr. Dennis Joseph. The computer funds for this work were provided by the Climate Dynamics Program in the Atmospheric Sciences Division of the National Science Foundation under Grant ATM78-07369.

1. Introduction

Since 1954, operational numerical weather prediction (NWP) has grown in sophistication and accuracy which has paralled the growth and development of computer technology. The first operationally useful NWP forecasts were issued in the late 1950's from the early beginnnings of the National Meteorological Center (NMC). The first numerical forecasts were based on a barotropic model covering most to the Northern Hemisphere (Cressman, 1958). In 1962, a three-level baroclinic model (Cressman, 1963) based on the primitive equations was put into operational use. Since 1966, a hemispheric six-layer baroclinic model (Shuman and Hovermale, 1968) was developed. A seventh layer was added in 1976. That model, together with the barotropic model and a regional Limited-Area Fine Mesh model (Howcraft, 1971) which was added in 1971, make up the NWP guidance used today. Operational forecasts have normally been issued out to 48 or 72 hours and as long as 84 hours in the case of the six-layer baroclinic model. Over the past 25 years, improvement has been substantial.

Skill of NWP model forecasts has been measured in many ways, but the earliest and still on-going measurement of skill is NMC's S1 score (Teweles and Wobus, 1954). This score is a measure of the error in the pressure gradient between standard grid points; a lower score is indicative of greater skill. The reader can obtain an appreciation of how much improvement has been made since the early stages of NWP from an inspection of the S1 score shown

in Fig. 1 (from Fawcett, 1977).

Improvement of NWP model forecasts can be attributed to a number of reasons. The most obvious of these are: a better understanding of the numerics, the basic dynamics, and the physics included in all models; the development of faster, more sophisticated computers allowing a greater space and time resolution; the ability to more accurately specify the initial state of the atmosphere through development of objective analysis routines capable of including both conventional and new types of meteorological information; and through detection of model deficiencies found by verification programs.

The primary purpose of any model verfication scheme is to document and evaluate the model performance characteristics with hopes of better understanding the strengths and weaknesses of physical and dynamical treatments incorporated into the models. Fawcett (1969) undertook a study to determine the systematic errors in the 36 hour PE at the 500 mb level. One result of Fawcett's study was to correct the negative bias over the Canadian Rocky Mountains. This error was diagnosed as a result of overly smoothed mountains in Western North America. After higher mountains were introduced into the PE in September, 1968 (Weather Analysis and Prediction Division, 1969), the negative bias was reduced significantly. Shown in Fig. 2 [from Fawcett (1969) Figs. 5b and 8c] are the mean daily error patterns for the months of October, 1965 and October, 1968. In another study of the PE, Leary (1970)

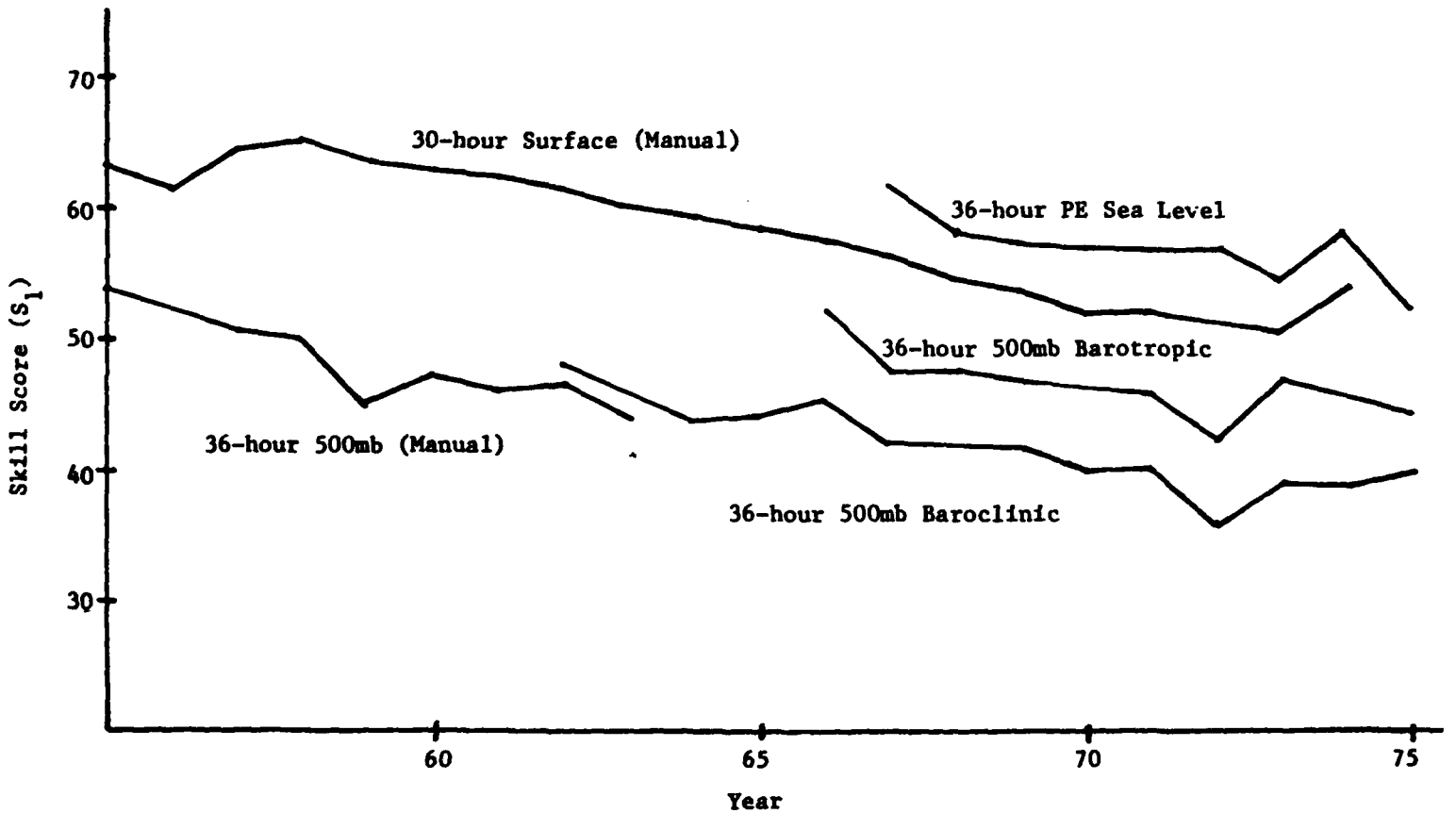
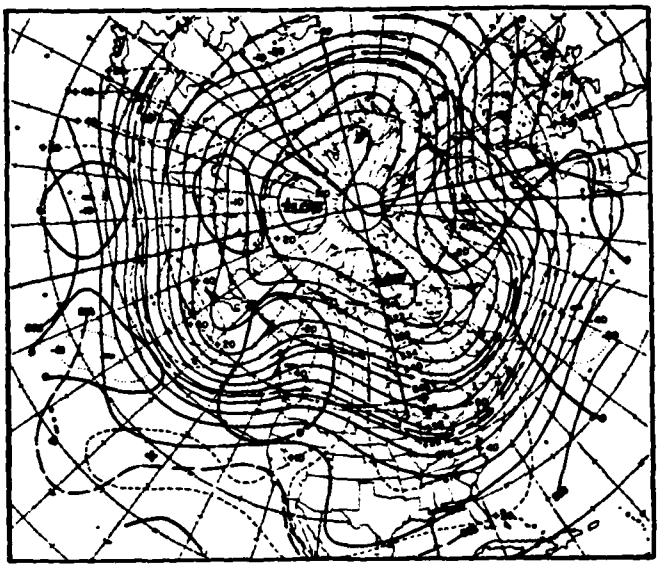
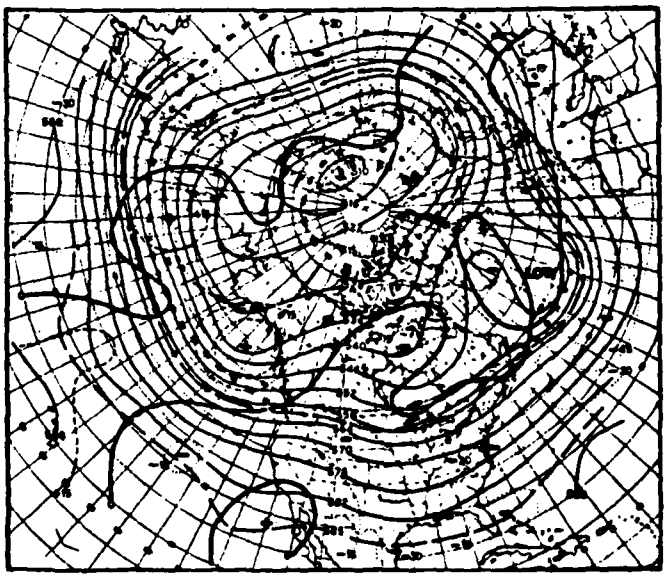


Fig. 1. Yearly mean S_1 scores for NMC's 36-hour sea-level prognoses from the PE, model, 30-hour manual sea-level prognoses, 36-hour barotropic model 500mb prognoses, 36-hour 500mb prognoses from the three-layer baroclinic model 1962-66 and the six-layer PE model 1966-75, and 36-hour 500mb manual prognoses. (taken from Fawcett, 1977)



(a)



(b)

Fig. 2. Mean monthly observed 500mb contours in decameters (solid lines) with mean observed jet stream core positions(heavy arrows) and mean error patterns for (a) 36-hour three level baroclinic, October, 1965 and (b) 36-hour PE model, October, 1968.

documented the systematic errors at the 1000 mb level by compositing forecast errors of the central sea level pressure of cyclones by storm type. In this way, she was able to document characteristics of model simulations of storms in the developing, mature and decaying stages. Pertinent conclusions from her study were: the depth of oceanic cyclones is generally underforecast; the depth of cyclones forming on the lee slopes of the Rocky Mountains is generally overforecast; most forecasts of thickness over land are too warm; and many forecasts of thickness over the oceans are too cold. No attempt has been made to correct these biases found by Leary.

The purpose of this thesis is to undertake a study to determine the systematic bias in the winter season geopotential height field of NMC's six-layer primitive equation (PE) model. Although other researchers have documented the systematic bias in the PE model, none have computed the bias at all gird points for a long term. Patterns of geographical bias found over the hemisphere and related to characteristics of the mean state of the atmosphere. It should be noted that this study covers the period after the PE was improved on the basis of Fawcett's results. In addition to the determination of the systematic bias in the PE, several other statistics were computed in order to gain more information about the local distribution of forecast error.

2. Data Description

The observational data set used in this study are the final analyses of the 500 and 1000 mb geopotential height for the Northern Hemisphere made by NMC. These data were obtained from the archives of the National Center for Atmospheric Research (NCAR). There are two maps per day, 00 and 12Z, over six winter seasons beginning with 1973-74 and ending with the 1978-79 winter season. A winter season is defined as December, January, and February. The forecast data set is comprised of 24-, 48-, and 72-hour 500 and 1000 mb forecasts of geopotential height over the same period and times as the observational data. The forecasts were made by NMC's six-layer PE developed by Shuman and Hovermale. A seven-layer PE was tested and became operational in January 1978. The forecast data were also obtained from NCAR. All statistics were calculated on the NMC octagonal grid containing 1977 grid points and results were mapped using the NCAR graphics software without any smoothing or interpolation to latitude/longitude grid points. The study is based on approximately 1030 24- and 48-hour 500 mb forecasts, each of which are equally divided between the 00 and 12Z initial times. There are about half as many, approximately 550, 72-hour forecasts which were made mostly from 00Z initial time. There were 48 missing grids almost equally divided amont the 00 and 12Z, 500 and 1000 mb categories.

3. Preliminary Comparison of Forecast Fields to Climatology

In order to establish a climatology for this study, the mean

and standard deviation of the observations were computed for both the 500 and 1000 mb levels. These maps are presented in Figs. 3 and 4. Comparison of the mean height of the 500 and 1000 mb levels for the six winter seasons in this study (Fig. 3) to climatology produced by other researchers [for instance, Blackmon (1976) Fig. 3 and Blackmon, et al. (1977), Fig. 1] shows good agreement in both position and amplitude of major standing wave troughs, ridges, highs, and lows. A similar comparison of the standard deviation fields for the two levels produced from this data set (Fig. 4) also shows good agreement in the position and amplitude of the maxima over the northern Atlantic and Pacific oceans but differs with respect to both position and amplitude of the maximum over central Asia.

As a preliminary comparison between forecast and observational climatologies, the mean and standard deviation for both the 500 and 1000 mb 72-hour forecasts are presented in Figs. 5 and 6, respectively. Comparison of Fig. 5a with the mean 500 mb observational climatology (Fig. 3a) shows good agreement with respect to the position of the mean troughs and ridges, but the mean 500 mb forecast height field tends to be higher in the polar regions and lower in the subtropics. At 1000 mb, the mean forecast climatology (Fig. 5b) is characterized by weaker lows in the northern Atlantic and Pacific oceans and a weaker Siberian high than the observed 1000 mb height field (Fig. 3b). The standard deviation of the 500 mb forecast field (Fig. 6a) appears quite similar to the

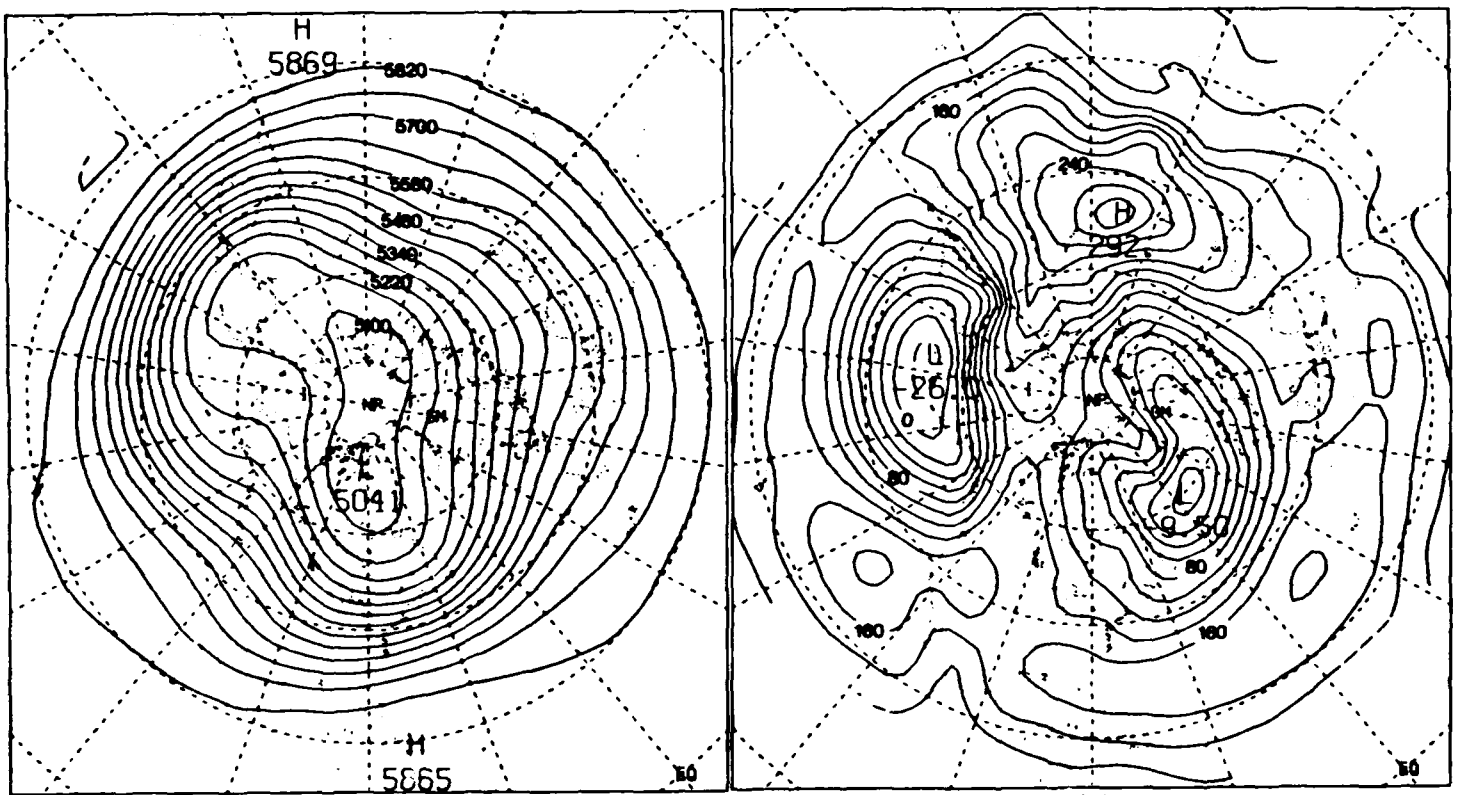
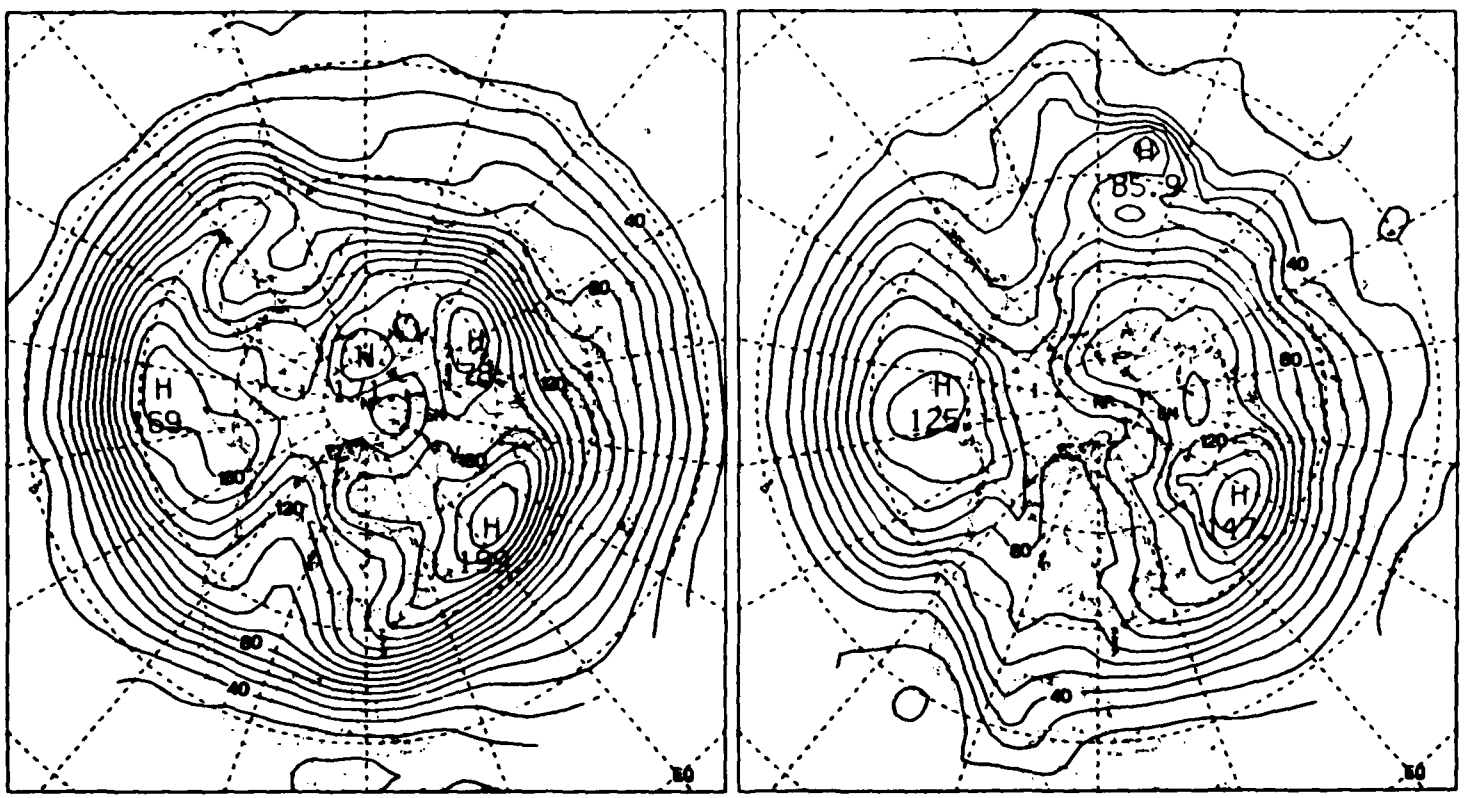
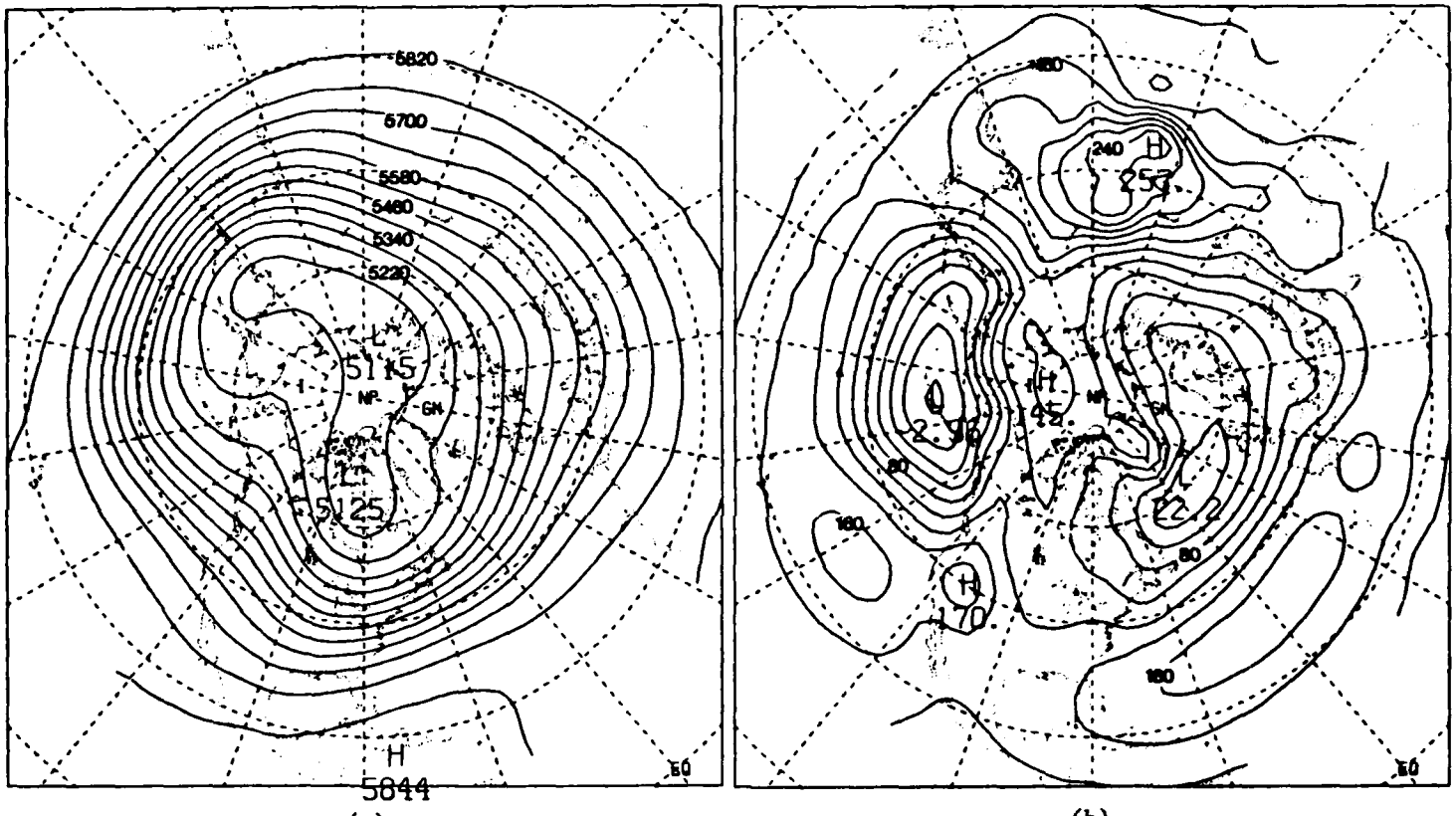


Fig. 3. Mean observed geopotential height at (a) 500mb and (b) 1000mb based on six winter seascns; contour interval for (a) 60m and for (b) 20m.



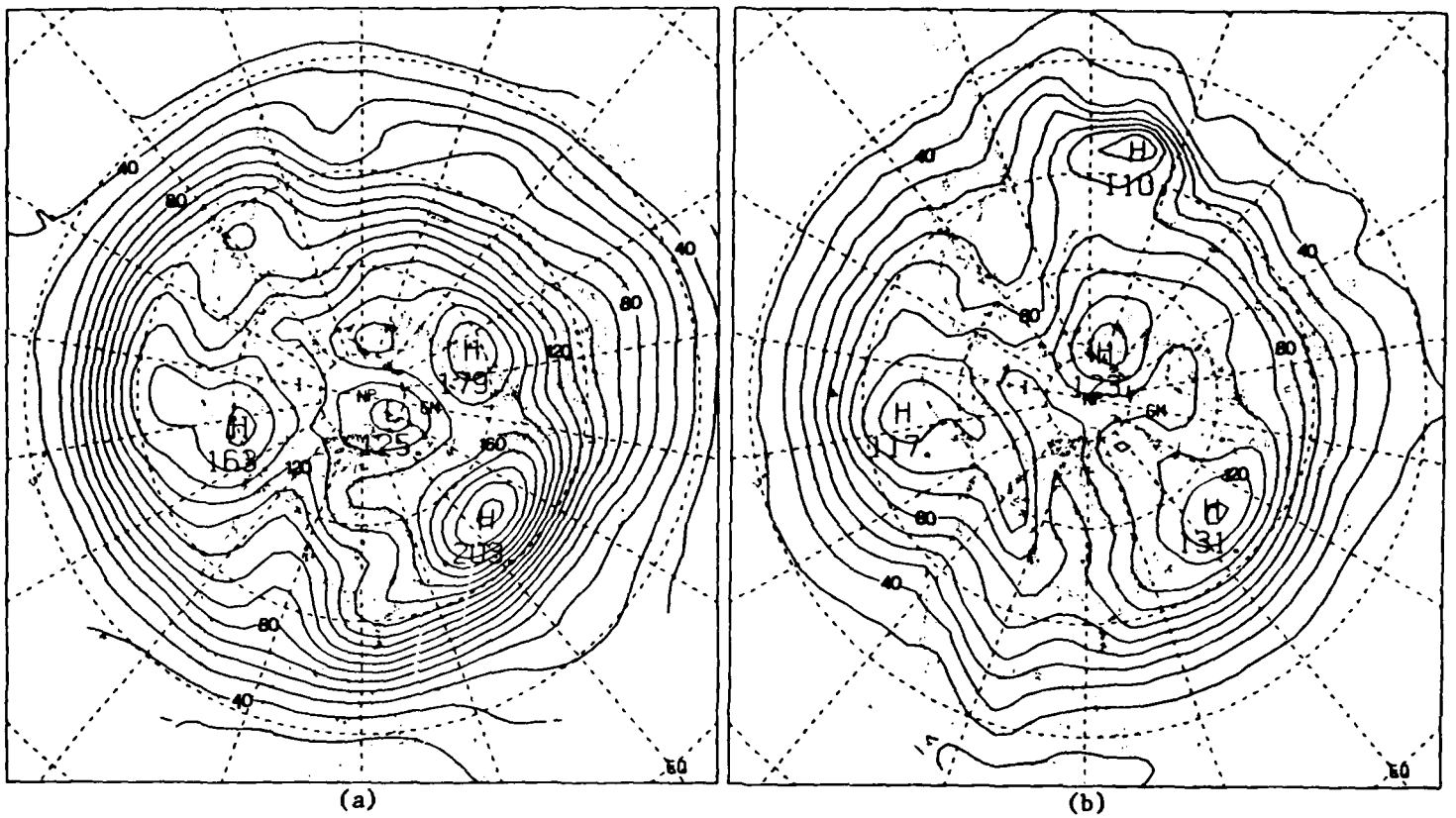
(a) (b)

Fig. 4. Standard deviation of the observed geopotential height for (a) 500mb and (b) 1000mb based on daily data for winter; contour interval 10m.



(a) (b)

Fig. 5. As in Fig. 3 except for 72-hour PE model forecast.



(a) (b)

Fig. 6. As in Fig. 4 except for 72-hour PE model forecast.

climatological one but slightly smaller values of the magnitude are observed. The standard deviation of the 1000 mb forecast (Fig. 6b) appears similar to that of the observation except that in the Tibetan plateau region, the model produces a variance approximately twice as large as the observed. Much if not all of the excess variance in this region may be associated with the problems of generating a model sea level pressure field in a region of high terrain. In summary, the forecast fields are characterized by weaker mean flows than the observed at both levels and they exhibit slightly less variability about the time mean. A more detailed discussion of the differences and determination of the systematic bias of the forecasts will be presented in Section 6.

4. Statistical Methods

A summary and discussion of currently used verification techniques and the organizations which actively participate in model verification programs can be found in Weather Forecasting and Weather Forecasts: Models, Systems, and Users, Summer Colloquium, 1976, NCAR, Boulder, Colorado. The most common objective verification statistics used by operational forecast centers (such as the U. S. Air Force Global Weather Center and NMC) are the mean algebraic error, the mean absolute error, the standard deviation of the error, NMC's S1 score, the root-mean-square-error, and the correlation coefficient between the forecast and observation. These are the so-called "single valued" statistics.

Frequently, more sophisticated objective verification statistics and techniques are used but primarily on research models. Since the primary goal of this study is to determine the geographical distribution of the bias and the non-systematic errors, only the mean algebraic error or bias, and the standard deviation of the error were computed.

The mathematical definitions of the bias, the standard deviation of the error, and the root-mean square error are:

Forecast Error $$E_i = F_i - O_i \quad (1)$$

Bias $$(\bar{E}) = \sum_{i=1} (E_i)/N \quad (2)$$

Standard Deviation of the Error $$(\overline{E'^2})^{1/2} = [\sum_{i=1} (E_i - \bar{E})^2/N]^{1/2} \quad (3)$$

Root-Mean-Square Error (r.m.s.) $$(\overline{E^2})^{1/2} = [\sum_{i=1} (E_i^2)/N]^{1/2} \quad (4)$$

where E is the forecast, O is the observation, and $(\bar{E})$ is the temporal mean. The summations are taken over the six winter seasons.

The relationship between $(\bar{E})$ and $(\overline{E'^2})^{1/2}$ can be more clearly understood by examination of Fig. 7 which shows the statistics for a hypothetical grid point. The model shows some skill forecasting the non-systematic part of the error as evidence by the fact that $(\overline{E'^2})^{1/2} < (\overline{O'^2})^{1/2}$. Some small fraction of the r.m.s. error for this hypothetical grid point is due to the positive bias of the forecast

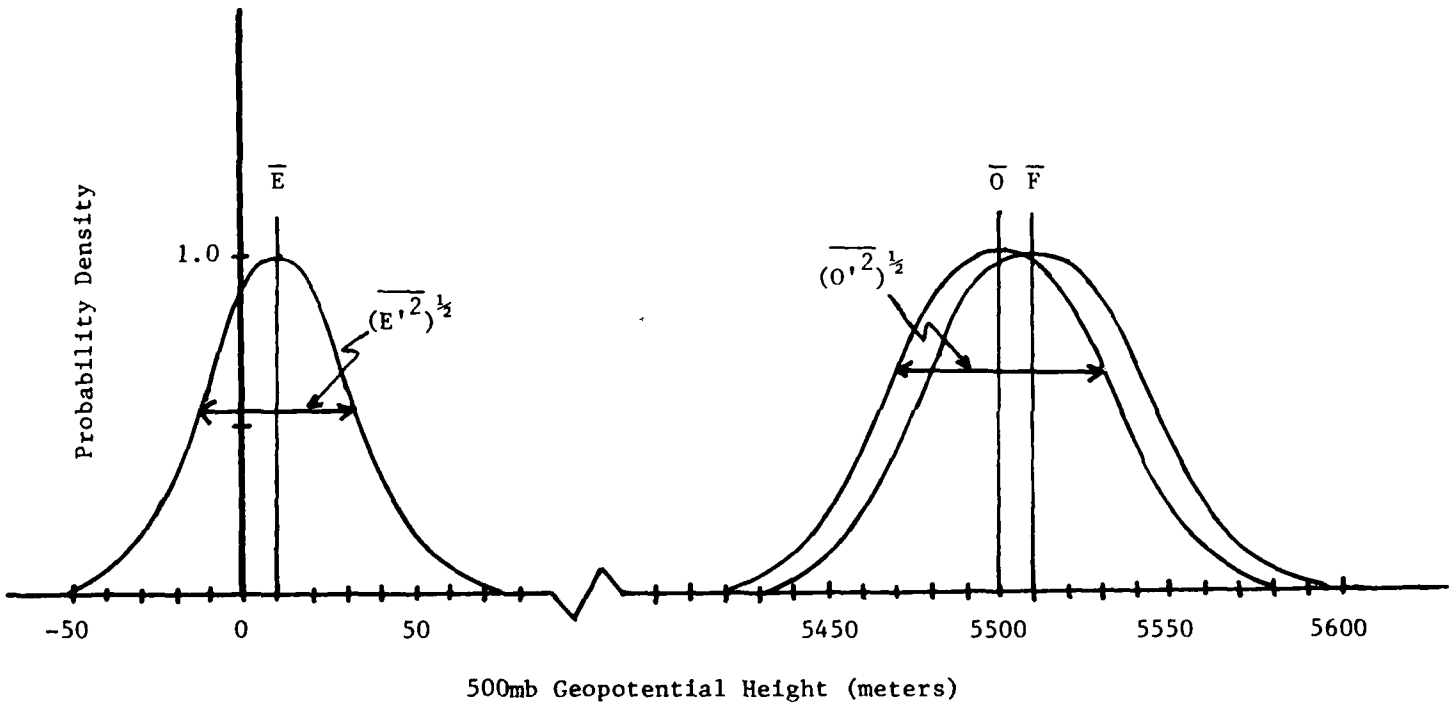


Fig. 7 Schematic relationship between the forecast, observation, and forecast error statistics for a particular grid point.

relative to the observation.

A measure of skill of a model is $(\overline{E'^2})^{\frac{1}{2}}$ or $(\overline{E^2})^{\frac{1}{2}}$ normalized by the climatological standard deviation of the variable in question. The first ratio is a measure of the non-systematic error of the model while the second ratio includes both the systematic and non-systematic components of the error. In either case, the lower the ratio, the more skillful the model. For a model with no skill for which $\overline{F} = \overline{O}$ and $\overline{F'^2} = \overline{O'^2}$, it can be shown that $\overline{E^2} = 2\overline{O'^2}$.

A qualitative graphical representation of the growth of the error as measured by the r.m.s. error, the theoretical limit of error growth in a perfect model, and the range of error growth of the current state of the art in NWP is shown in Fig. 8. Here the growth of the error in a perfect model is due only to the inherently limited predictability of the atmosphere and not to any error in the initial conditions or deficiencies in the model's representation of dynamical or physical processes. Many studies of forecast skill use this type of graph as one tool for comparison.

Another measure of skill is the correlation coefficient between the forecast and the observation. The correlation coefficient between these two variables can be computed using equation (5).

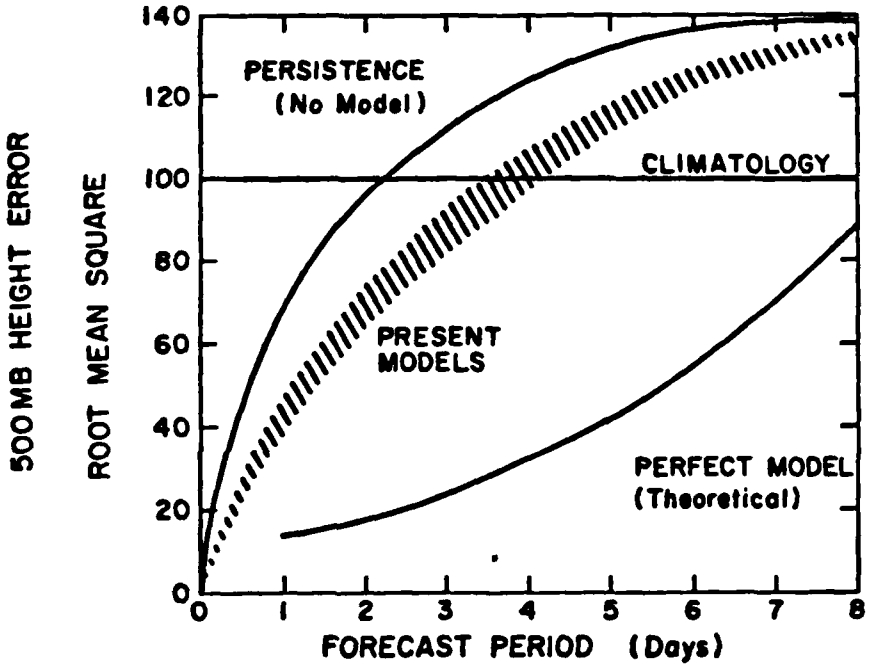


Fig. 8. Qualitative comparison of "perfect model" forecast skill with that of persistence, climatology, and present models, as measured by the r.m.s. 500mb height error (Leith,1976).

$$r(F,O) = \frac{\overline{F'O'}}{(\overline{F'^2})^{\frac{1}{2}}\,(\overline{O'^2})^{\frac{1}{2}}} = \frac{\overline{F'^2} + \overline{O'^2} - \overline{(F-O)'^2}}{2(\overline{F'^2})^{\frac{1}{2}}\,(\overline{O'^2})^{\frac{1}{2}}} \qquad (5)$$

The second form of the equation is convenient for determining r(F,O) because only the standard deviation of the forecast is required in addition to the already computed standard deviation of the forecast error and the climatological standard deviation.

By breaking up the total error into systematic and non-systematic parts, the percent of the total variance due to the bias can be determined.

$$\text{\% Variance due to bias} = \frac{\overline{E}^2}{\overline{E^2}} = \frac{\overline{E}^2}{\overline{E}^2 + \overline{E'^2}} \qquad (6)$$

This statistic gives an indication of the seriousness of the error associated with the bias. Where the bias contributes a significant part of the error, it might be advisable to correct the forecast for the bias at those grid points most seriously affected.

Additional information about the model characteristics can be gained by comparing the error statistics for the 500 and 1000 mb levels. The correlation between the error at the 2 levels and the ratio of the standard deviation of the errors at the two levels are two statistics which will give an indication of the vertical structure of the error field.

A comparison of the normalized standard deviation of the errors, $(\overline{E'^2}/\overline{O'^2})^{\frac{1}{2}}$ or $(\overline{E}^2/\overline{O'^2})^{\frac{1}{2}}$, between the two levels gives an

indication of whether the model is more skillful at 500 mb level than at the surface. Still other insights into the model characteristics can be gained by analyzing the time series of the forecast, observation, and forecast error at individual grid points. For example, the cross correlation of the forecast and observed time series gives information concerning systematic phase errors in the tests. All of these statistics and comparisons were performed and will be discussed in section 6.

An important source of uncertainty in this study is the assumption that the final NMC objective analysis perfectly matches the actual state of the atmosphere. It has been shown by Oort (1978) in his study of the adequacy of the rawinsonde network, that such analyses may contain significant errors. On the basis of experiments with a simulated network of radiosonde stations embedded in a General Circulation Model grid, he estimated that the root-mean-square error averaged over the month of January using daily values is on the order of 14m. It will be shown in the next section that some of the largest biases occur in regions of relatively dense radiosonde observations.

5. Investigation of Effects of Different Analysis Procedures

During the period of study, three different objective analysis schemes were operational at NMC: the Cressman routine (Cressman, 1959) which was operational until September 1974; the Flattery-Hough routine (NWS, Tech. Proc. Bull., 1974) which was used

from that time through September 1978, and the present routine, the Optimum Interpolation method (McPherson, 1979), was operational during the last winter season in this data set. To test the influence of the different analysis schemes on the statistics presented here, the mean error of the 500 mb 72-hour forecast was computed for the periods in which each of the three analysis routines were in use. The results for the periods during which the Flattery-Hough and the Optimum Interpolation routines were used are shown in Fig. 9. If one bears in mind that the Flattery-Hough statistics are based on four winter seasons of forecasts and the Optimum Interpolation statistics are based on only one, the patterns are perhaps as similar as one can expect. Thus it is concluded that the objective analysis schemes had little effect on the systematic errors of the model.

6. Error Statistics

a. Systematic Errors

Shown in Fig. 10 are mean error patterns of the PE at the 500 mb level. At 24 hours, a negative bias is present over low latitude oceans and a generally positive bias poleward of 40°N. By 72 hours, there is a stronger longitudinal dependence of the bias with positive bias in the mean troughs and over the polar region, and negative bias in the mean ridges and low latitudes. The bias over the pole increases by about 20 meters per 24 hours, which is typical of the largest errors at this level.

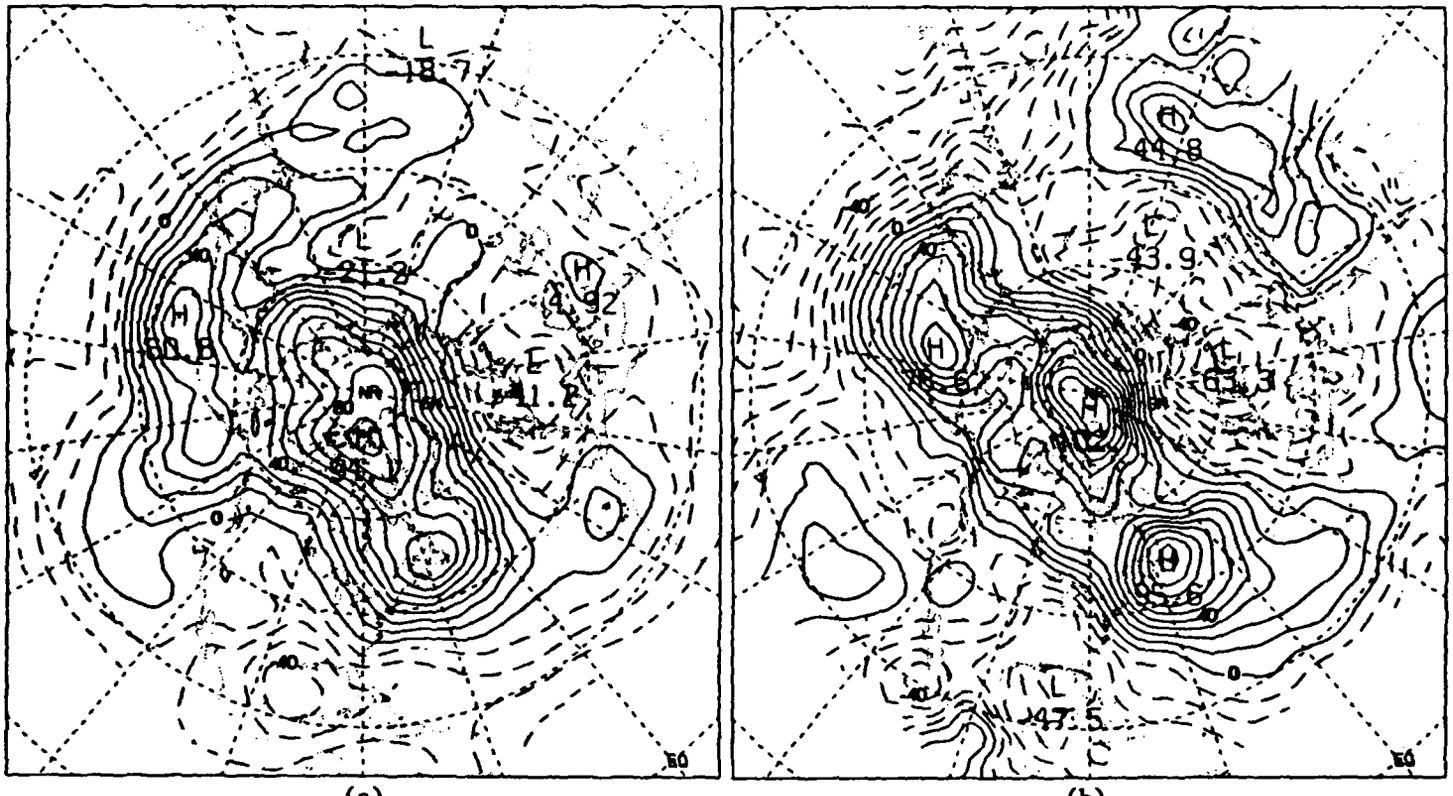


Fig. 9. Mean error for the 500mb 72-hour forecast for the period in which the (a) Flattery-Hough and (b) Optimum Interpolation objective analysis schemes were operational; contour interval 10m.

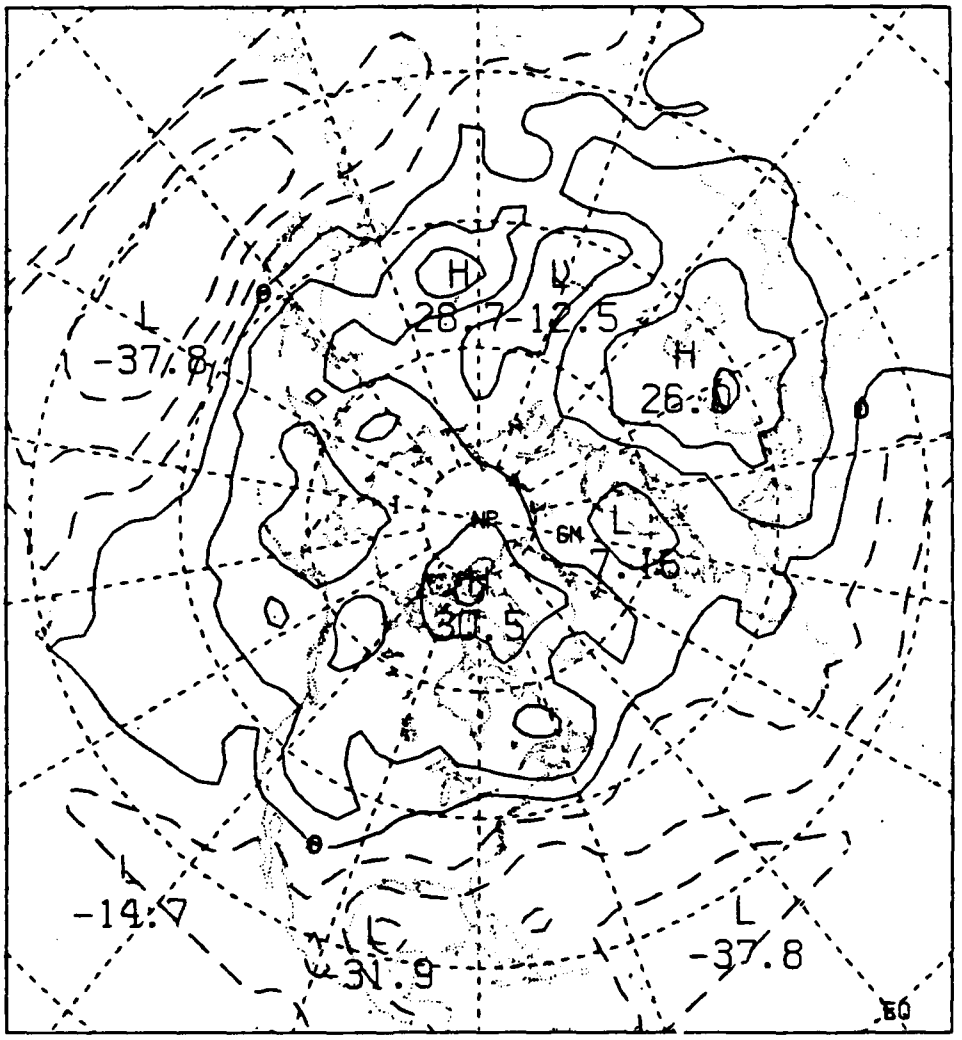


Fig. 10a. Mean error for the 500mb 24-hour forecast; contour interval 10m.

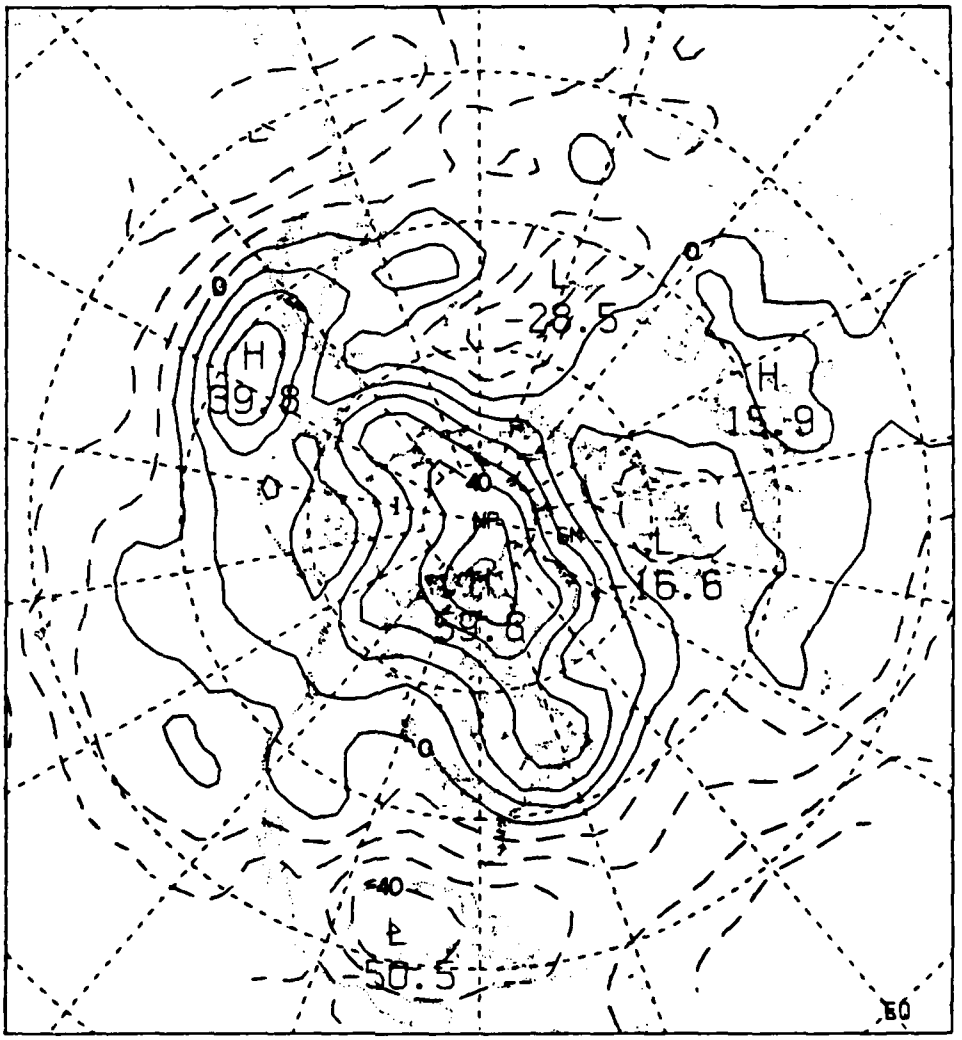


Fig. 10b. As in Fig. 10a except for 48-hour forecast.

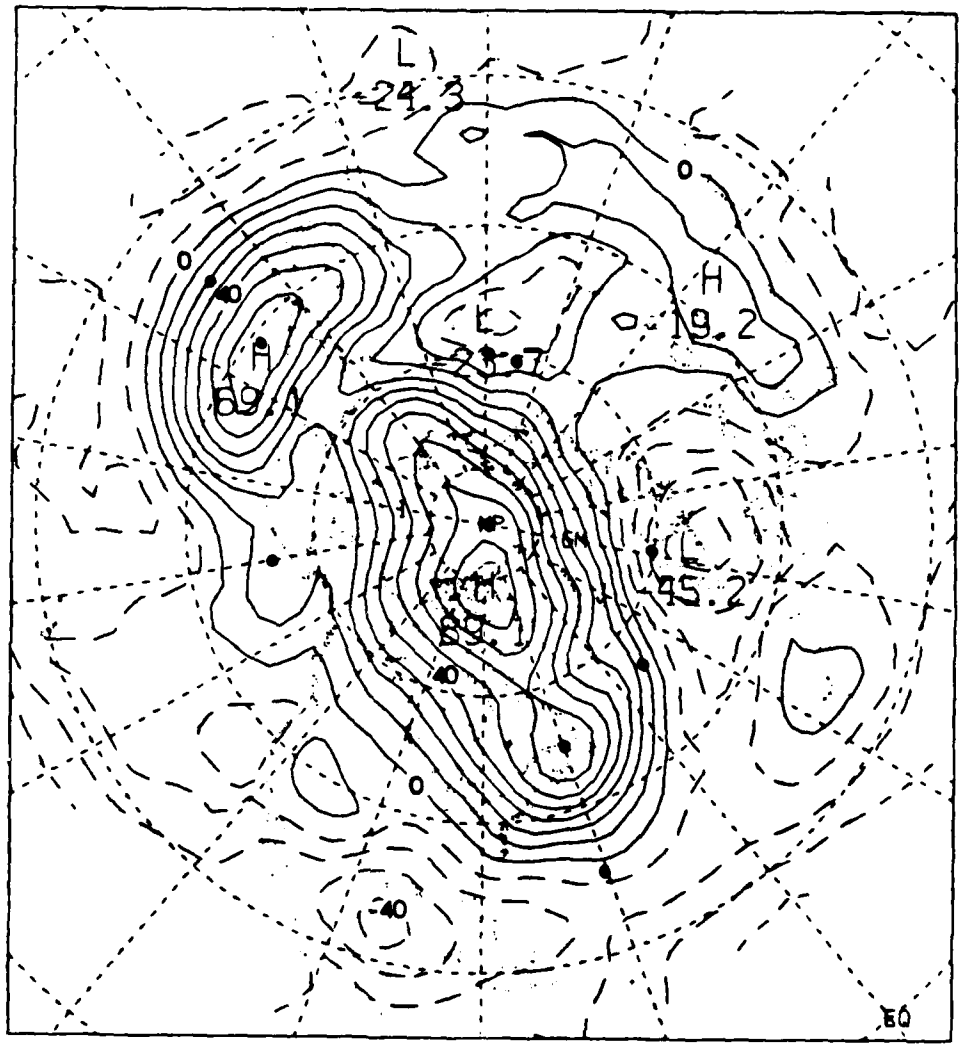


Fig. 10c. As in Fig. 10a except for 72-hour forecast. Dots indicate grid points at which cross spectrum analysis was performed.

Shown in Fig. 11 are the corresponding mean error patterns at the 1000 mb level. In contrast to the 500 mb height error, the 1000 mb height errors do not show as much of a latitudinal gradient of the bias. The bias pattern is generally oriented toward negative bias over continents, and slightly negative bias over the low latitude oceans. By 72 hours, centers of positive and negative bias closely match the positions of the climatological mean wintertime lows and highs.

The 1000 mb bias fields (Fig. 11) reveals some features which may be associated with major mountain ranges. There tends to be a west-to-east, positive-to-negative bias across the mountain ranges. For example, the 24 hour forecast error field shows a center of negative bias located in the Gulf of Mexico. At 48 hours it begins to expand north along the east slope of the Rocky Mountains, and becomes more negative at 72 hours. At the same time, a ridge of positive bias develops over and to the west of the range. Other mountain ranges where this cross range bias is seen are over the Himalayas and the Alps.

The thickness error was computed from the mean 500 and 1000 mb error fields using the relation,

$$(\overline{F-O})_{500} - (\overline{F-O})_{1000} = (\overline{F}_{500}-\overline{F}_{1000}) - (\overline{O}_{500}-\overline{O}_{1000}). \qquad (7)$$

Fig. 12 shows the thickness error pattern for the 72-hour forecast. It is particularly obvious that the mean 500 mb troughs are too warm and the mean ridges are too cold.

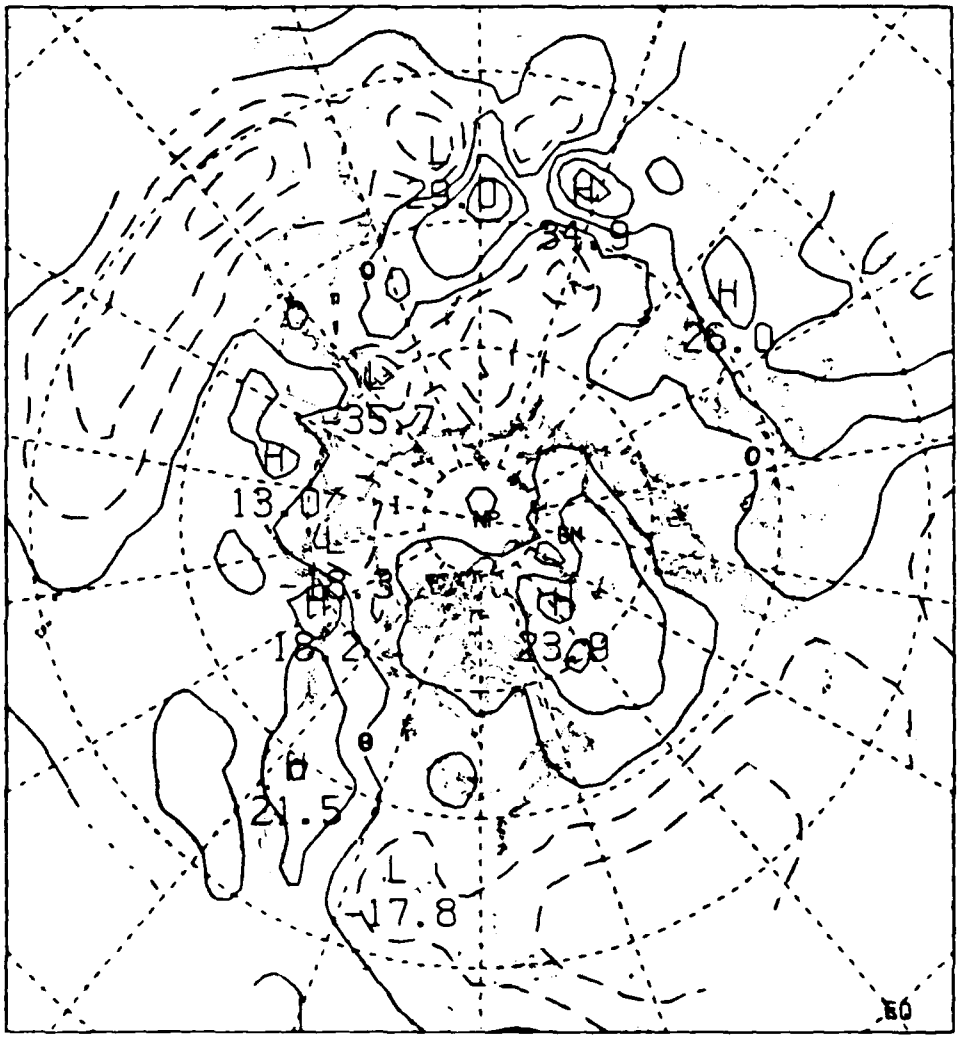


Fig. 11a. Mean error for the 1000mb 24-hour forecast; contour interval 10m

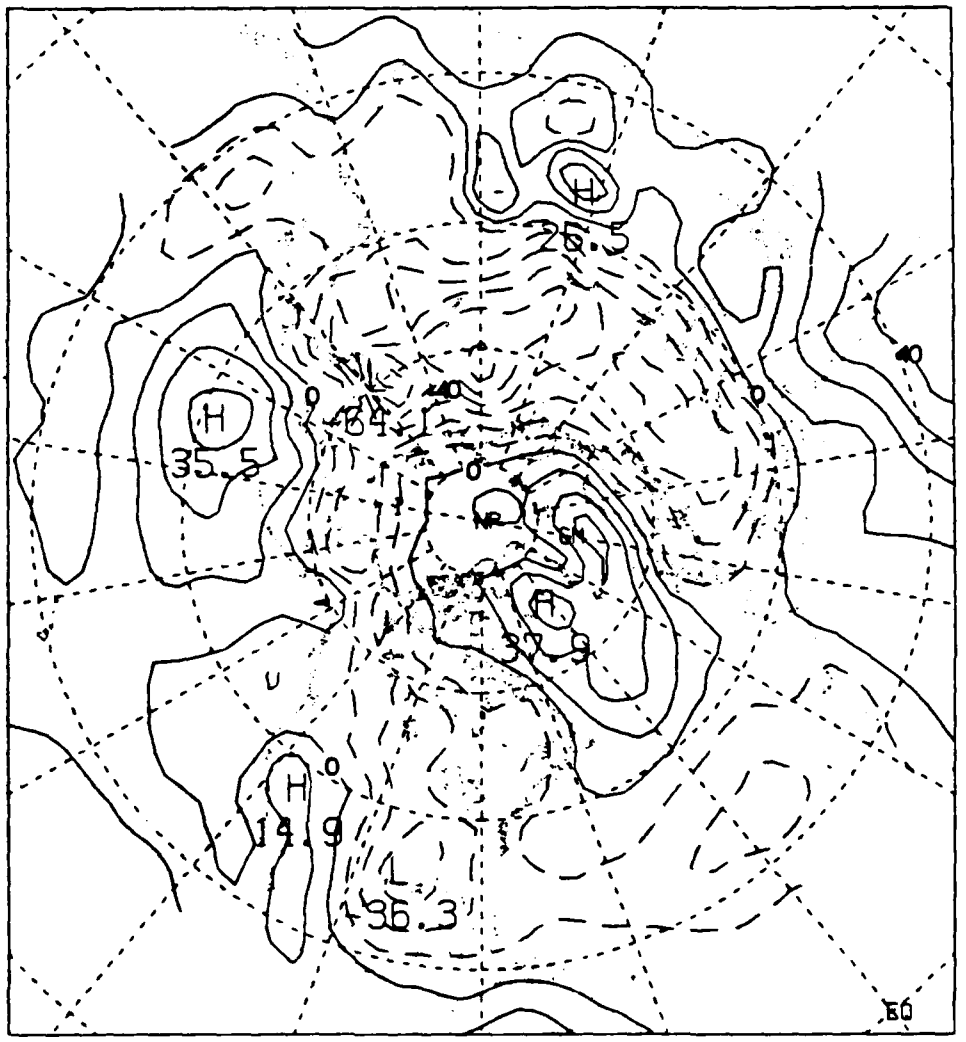


Fig. 11b. As in Fig. 11a except for 48-hour forecast.

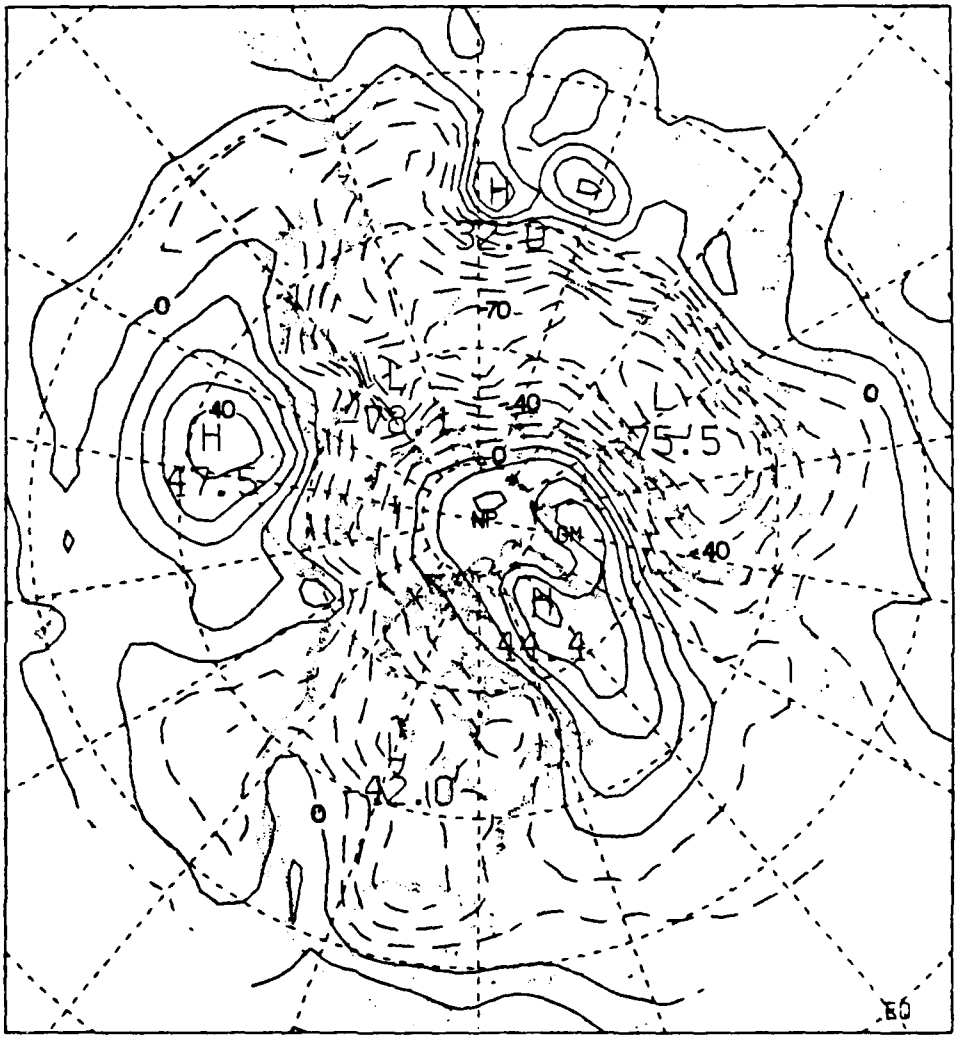


Fig. 11c. As in Fig. 11a except for 72-hour forecast.

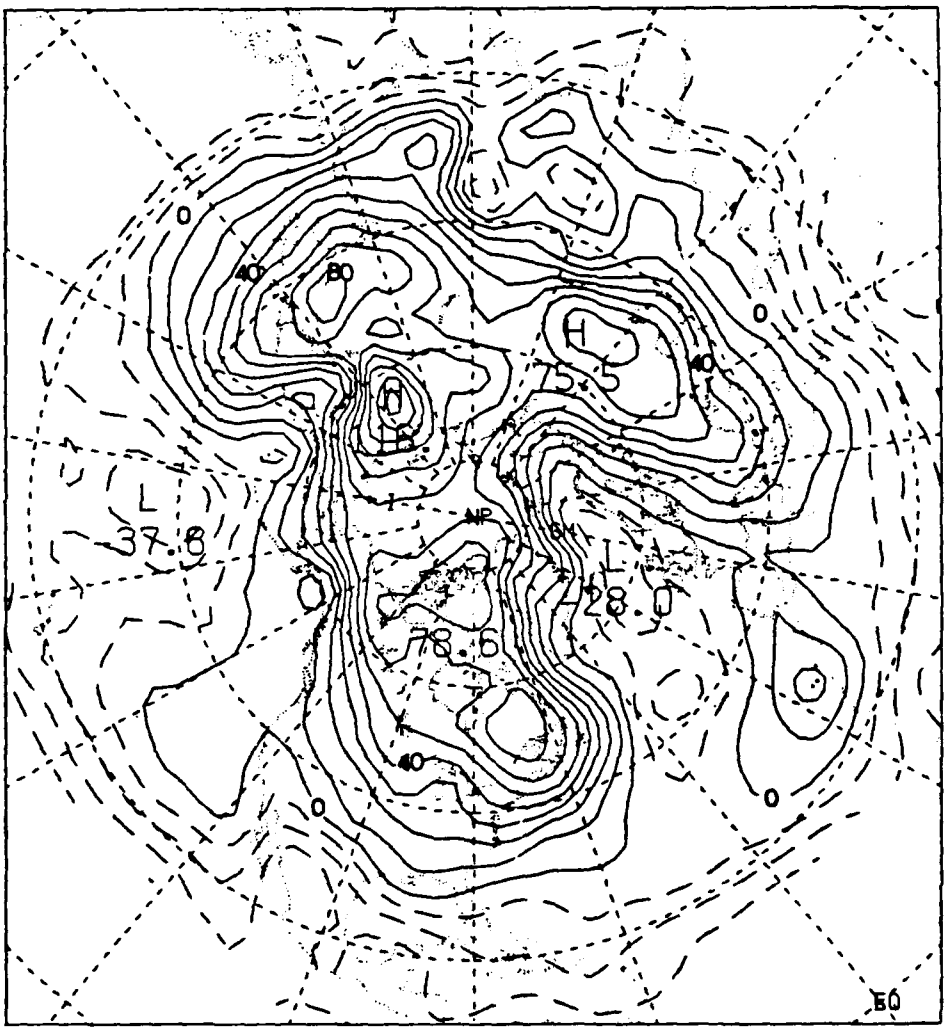


Fig. 12. Mean 500-1000mb thickness error for the 72-hour forecast; contour interval 10m.

It is evident from these results that the model systematically damps out the climatological mean surface highs and lows, and 500 mb troughs and ridges. This behavior has been detected previously by other researchers. Fawcett (1969) in his study of the systematic error in the 36-hour 500 mb PE model forecast (1964-1968) detected the tendency for the model to flatten out the fields not only in mean monthly charts, but also in daily error charts. Baumhefner (1978) noted the growth in the positive bias at higher latitudes as the forecast period increased which is a tendency consistent with the growth of the positive bias over the pole obtained in the present study. It is interesting to note that systematic damping of the amplitude of the atmospheric fields also has been observed to occur in individual disturbances. For example, Baumhefner (1978) also noted that the PE underforecasted the amplitude of the large scale waves on a day-to-day basis and at all forecast times.

b. Non-Systematic Errors

Shown in Fig. 13 is the ratio, $(\overline{F'^2}/\overline{O'^2})$ Over most of the mid-latitudes poleward of 30°N, this ratio is close to unity. Even though it was demonstrated that the PE model shows a tendency to systematically underforecast climatological features as well as individual disturbances, the closeness of this ratio to one indicates that the PE model is able to simulate real atmospheric fluctuations more accurately than expected.

Shown in Fig. 14 and 15 are the patterns of standard deviation

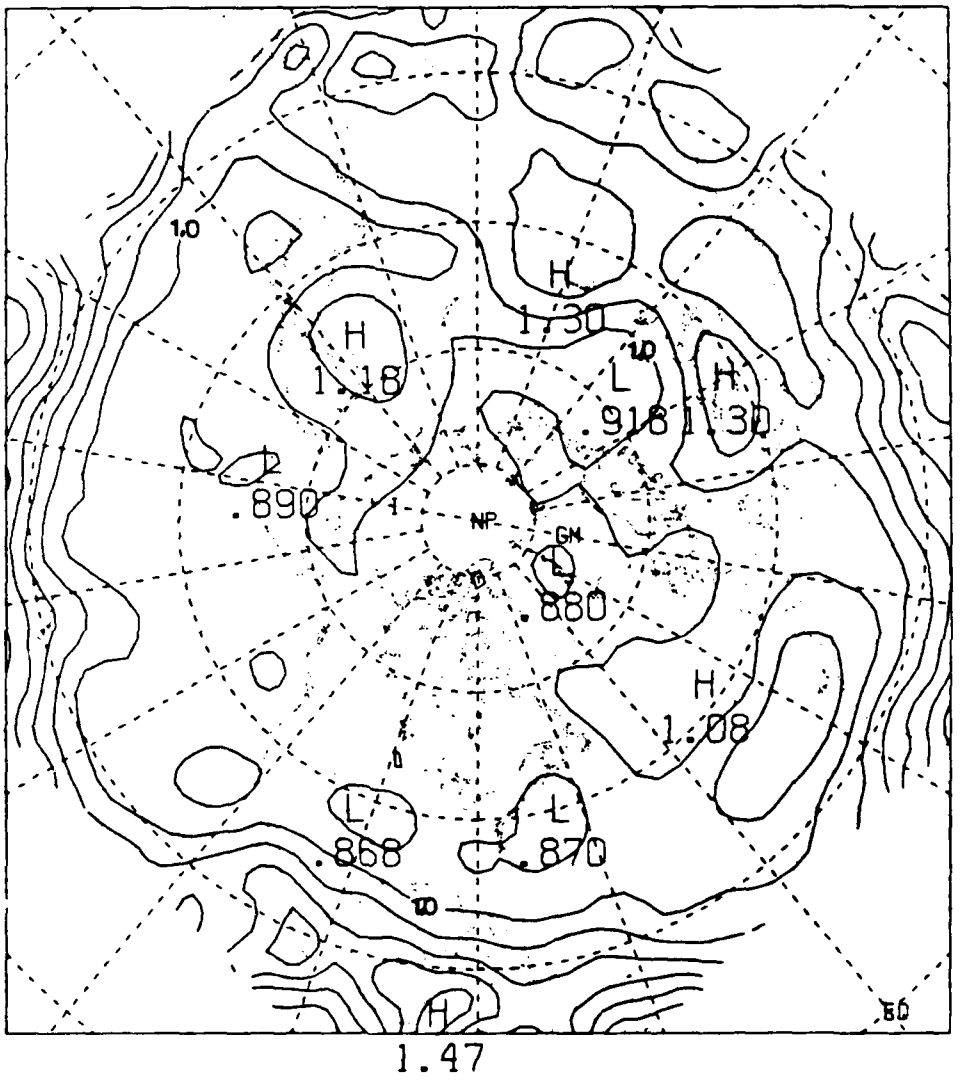


Fig. 13. Ratio of the standard deviation of the 500mb 72-hour forecast to the 500mb climatological standard deviation of the observation; contour interval 0.1.

of the error at the 500 and 1000 mb levels respectively. At both levels, it can be seen that there is a monotonic increase in the non-systematic error over the forecast period at almost all grid points. Also, the shape of the standard deviation patterns clearly resemble the climatological standard deviation as can be seen by comparison with Fig. 4. They even more strongly resemble the bandpass filtered patterns shown in Fig. 5d, [Blackmon (1976)], and Fig. 2b, [Blackmon, et al. (1977)]. In regions with large day-to-day variability, such as the mid-latitude storm tracks, the standard deviation of the error is relatively large while in regions with small day-to-day variability, such as the tropics, the standard deviation of the error is relatively small.

Shown in Fig. 16 is geographical distribution of the normalized standard deviation of the error, $(\overline{E'^2}/\overline{O'^2})^{\frac{1}{2}}$ for both the 500 and 1000 mb 72-hour forecasts. Note that the normalized fields are relatively featureless except for the monotonic increase as one approaches the equatorward edge of the grid.

The correlation coefficient between the 500 mb 72-hour forecast and the verifying observation, as computed from Eq. 5, was used as another measure of forecast skill. Contours of correlation coefficients plotted at each grid point are shown in Fig. 17. Comparison to the normalized standard deviation field (Fig. 16a) shows the same general shape with duplication of all major features. Note the same monotonic pole-to-equator decrease in skill of the forecast.

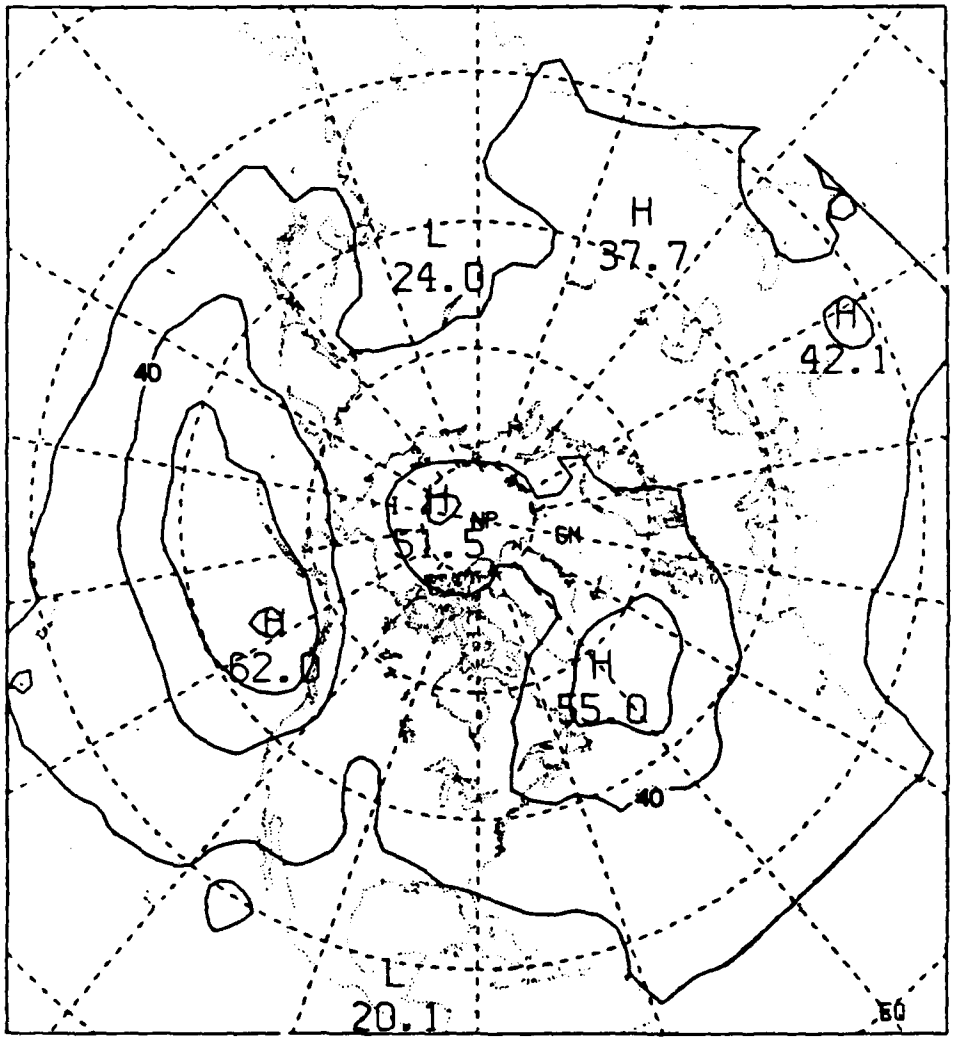


Fig. 14a. Standard deviation of the 500mb 24-hour forecast error; contour interval 10m.

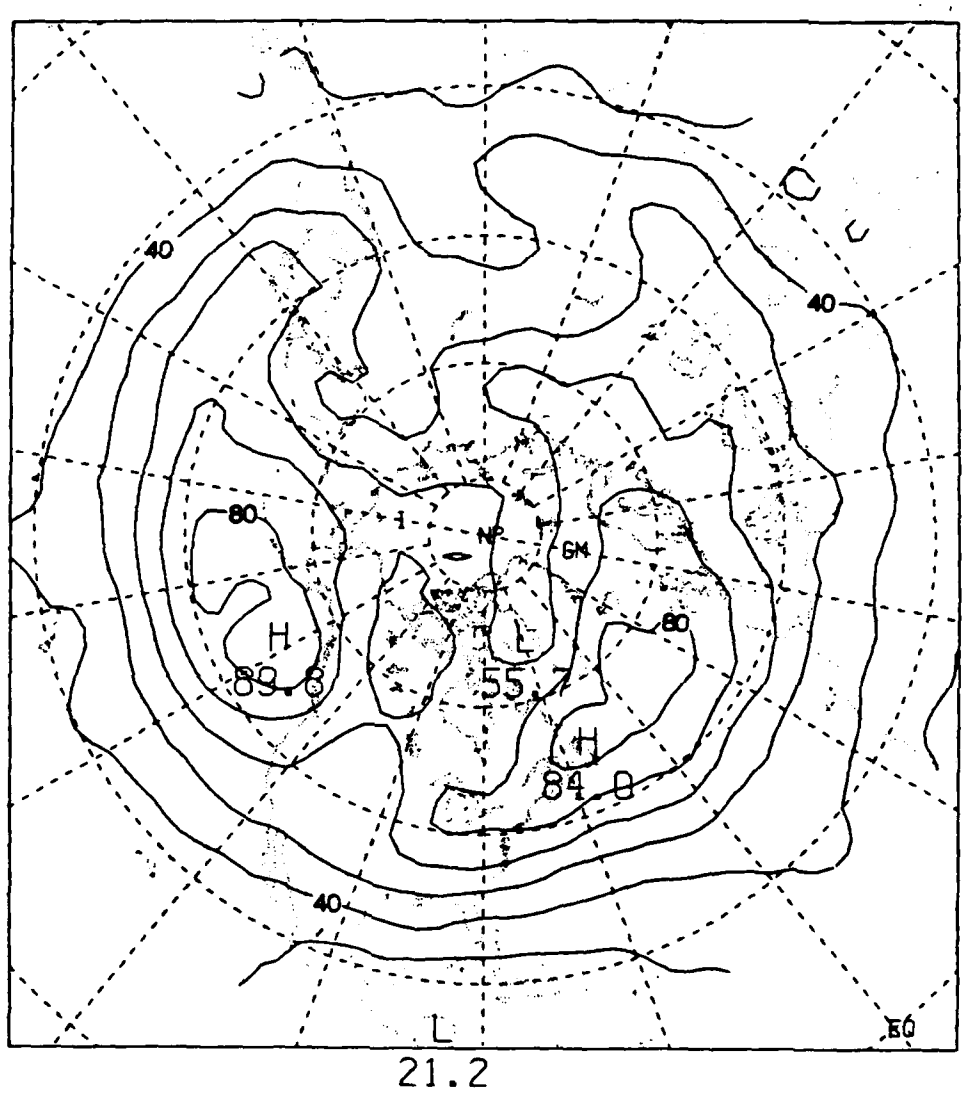


Fig. 14b. As in Fig. 14a except for 48-hour forecast.

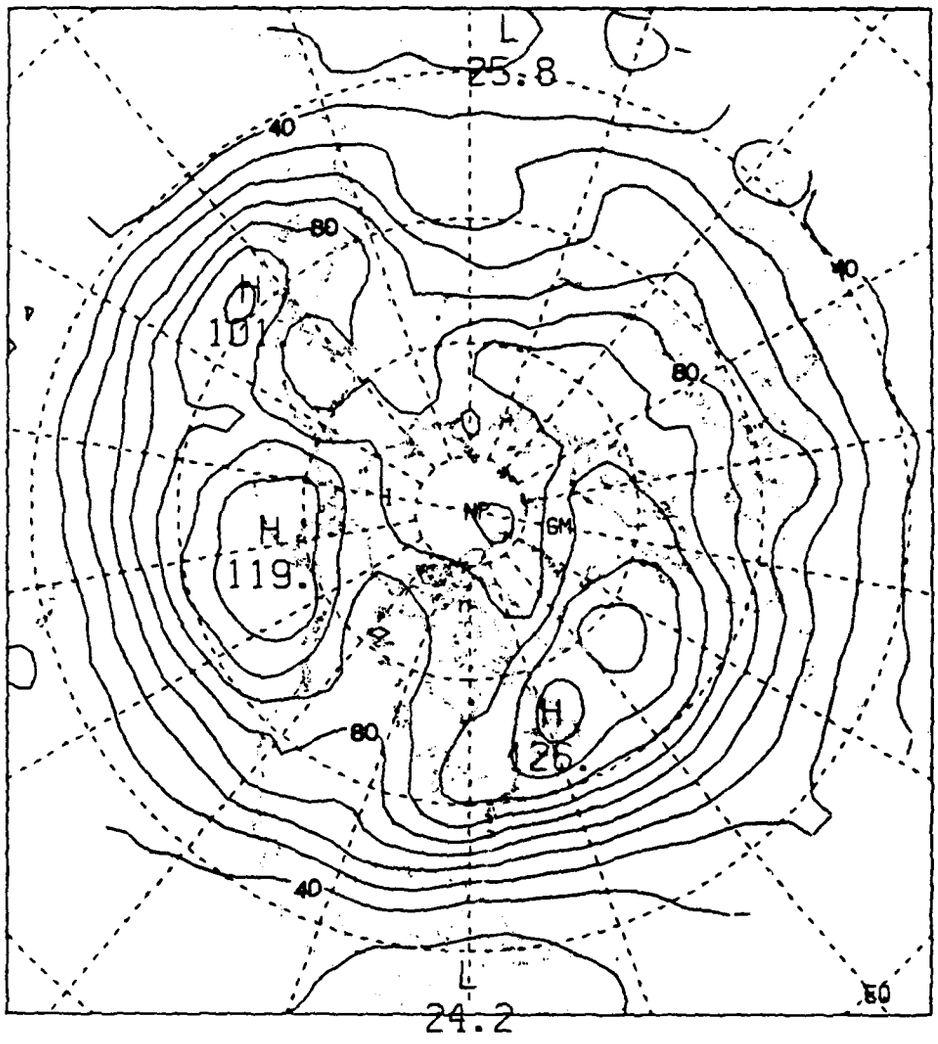


Fig 14c. As in Fig. 14a except for 72-hour forecast.

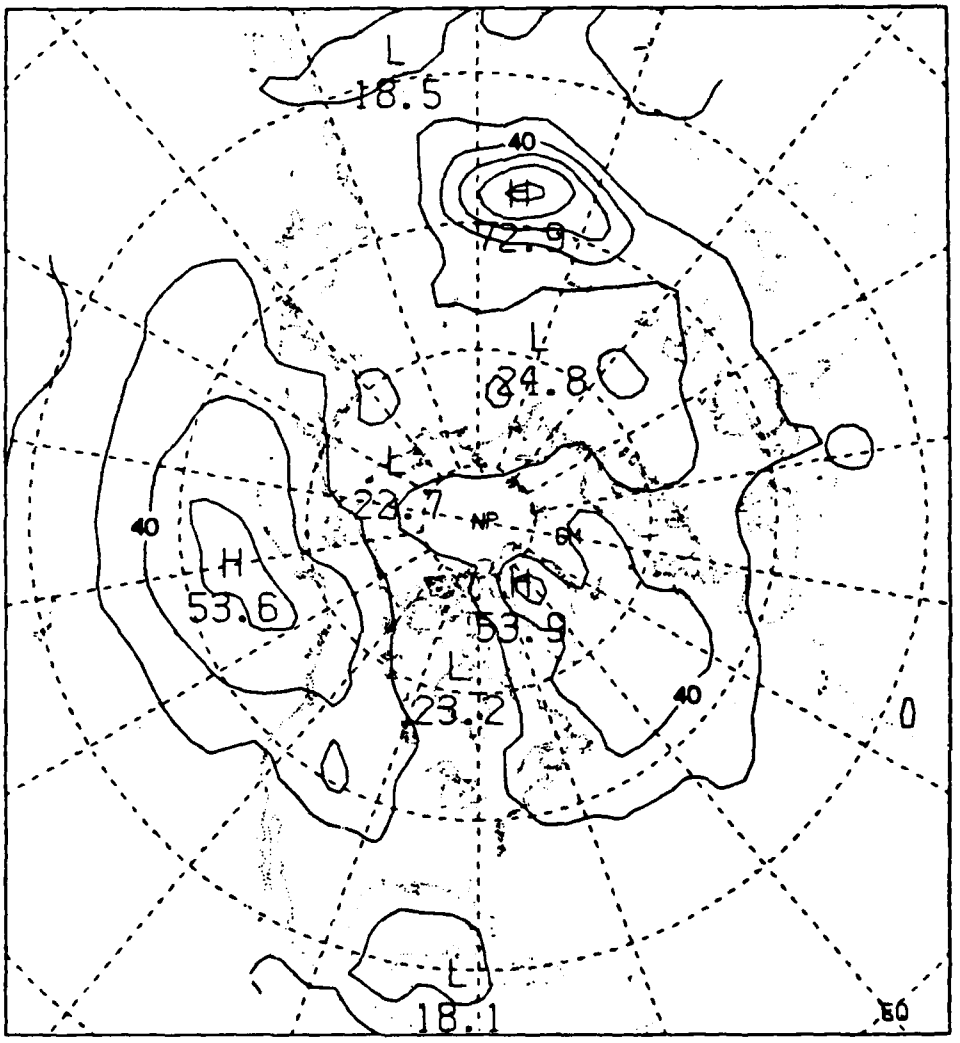


Fig. 15a. Standard deviation of the 1000mb 24-hour forecast error; contour interval 10m.

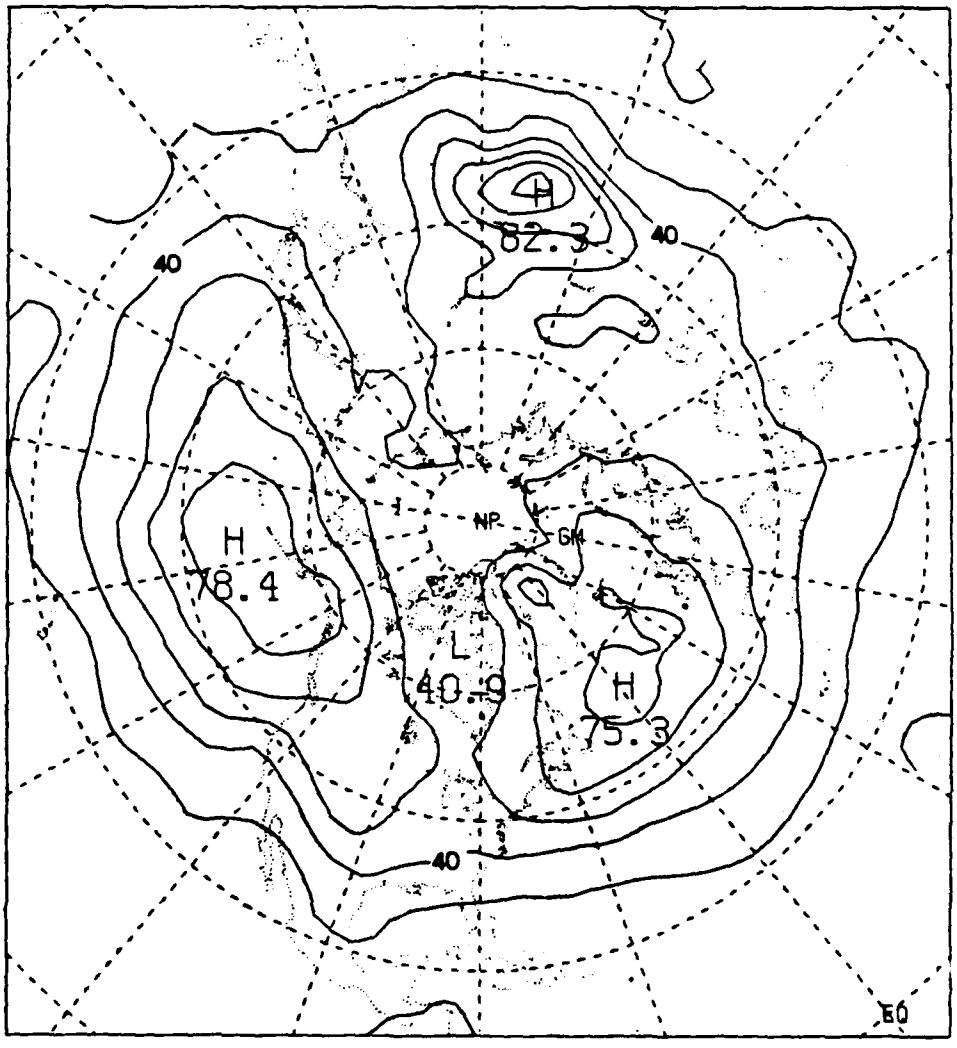


Fig. 15b. As in Fig. 15a except for 48-hour forecast.

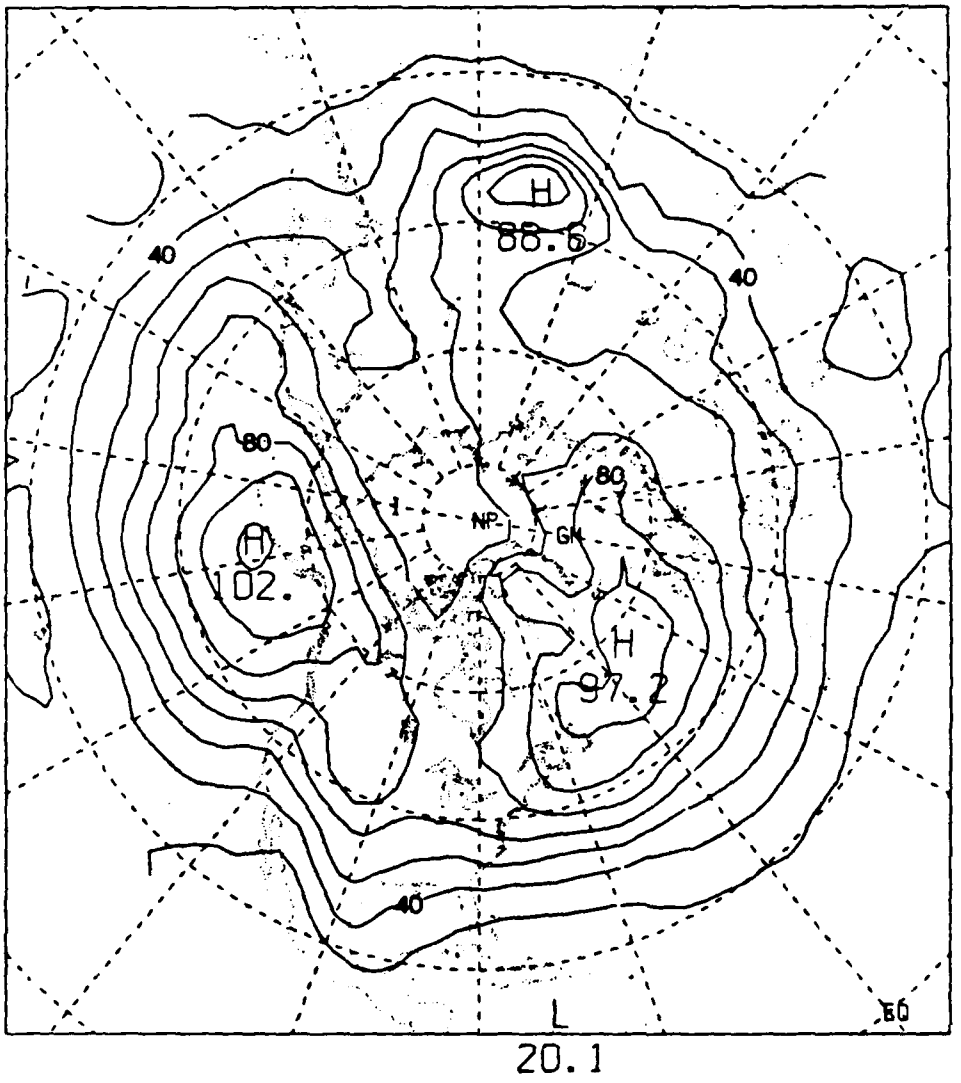


Fig. 15c. As in Fig. 15a except for 72-hour forecast.

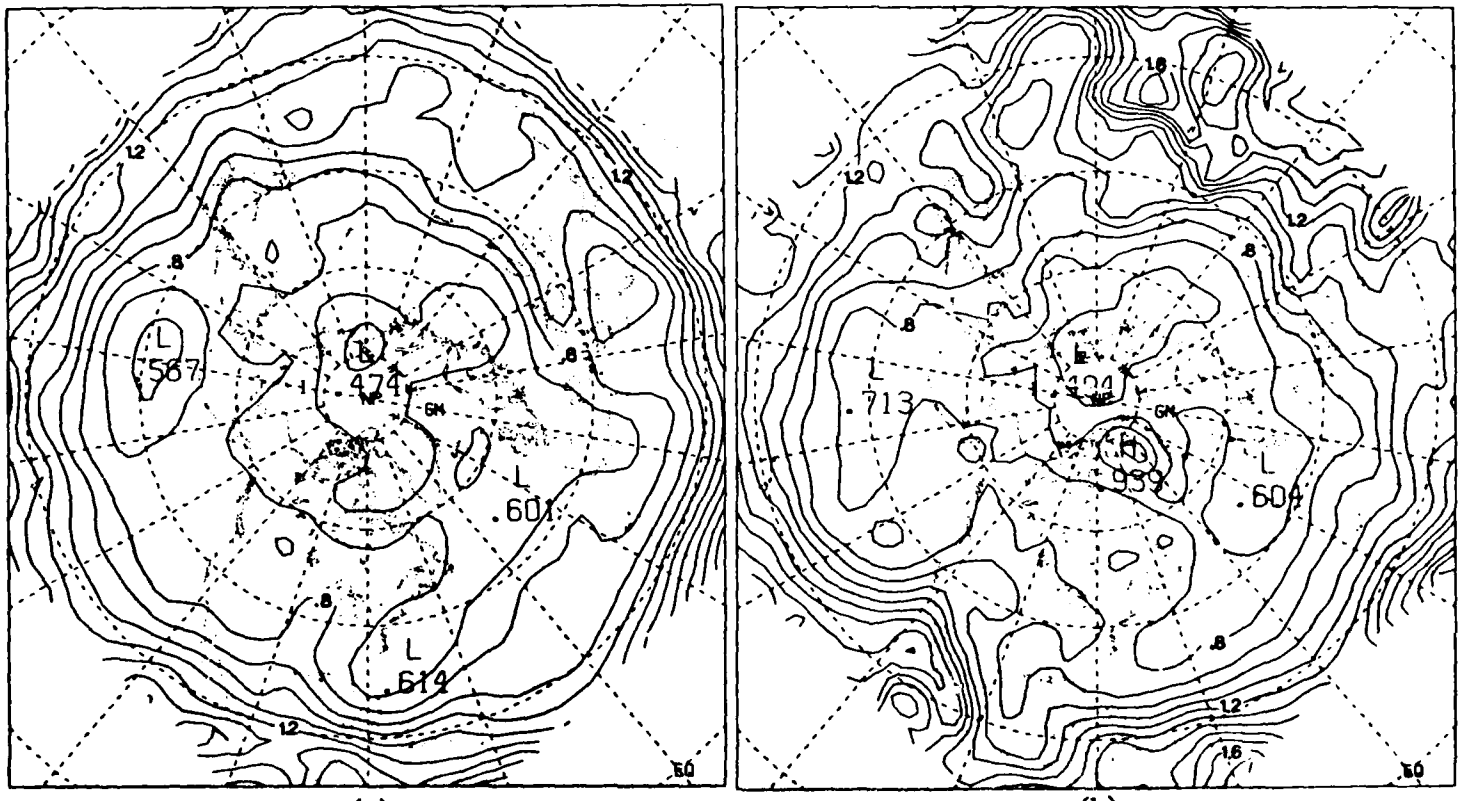


Fig. 16. Normalized standard deviation of the error for the (a) 500mb and (b) 1000mb 72-hour forecasts; contour interval 0.1.

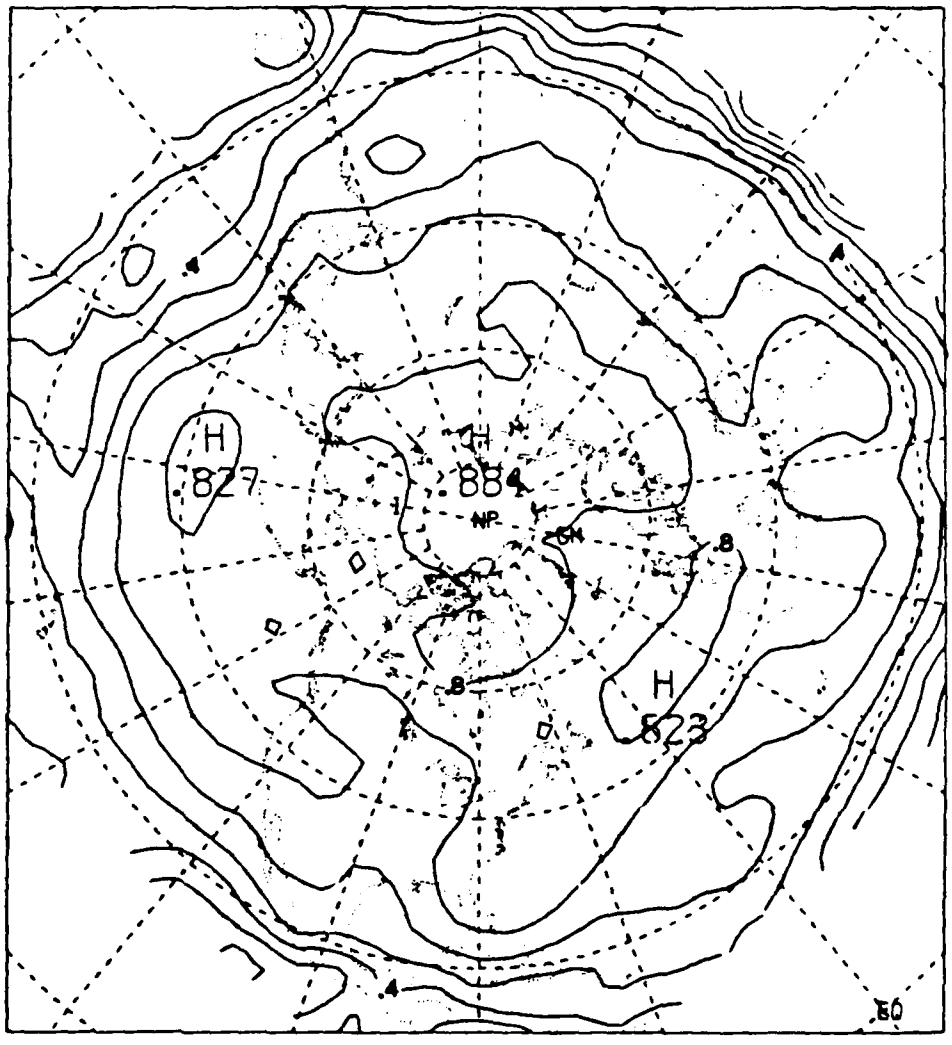


Fig. 17. Correlation coefficients between the forecast and observation for the 500mb 72-hour forecast; contour interval 0.1.

In order to gain an appreciation of the growth of the error at both the 500 and 1000 mb levels, graphs of the area-averaged standard deviation of the error, non-systematic error, and total error are plotted versus forecast time in Fig. 18. The area-averaged climatological standard deviation derived from this data set is plotted as a horizontal line. At both levels, the PE shows skill relative to climatology out to at least 72 hours. The 500 mb errors are larger in an absolute sense, but when normalized by dividing through by the climatological standard deviation, they are smaller than the corresponding 1000 mb errors.

The percent of the total variance of the error due to the bias, $\overline{E}^2/(\overline{E}^2 + \overline{E'^2})$, was calculated for the 500 mb 72-hour forecast and is plotted in Fig. 19. Most of the error is due to the non-systematic part of the total error as indicated by the low values, but the bias contributes a significant part of the total error over the pole, at low latitudes, and in the mean troughs.

c. Vertical Structure of the Error

Various statistics were calculated and compared between the 500 and 1000 mb levels in order to determine the vertical structure of the error field. The ratio of the 500 to 1000 mb standard deviation of the error, $(\overline{E'^2}_{500}/\overline{E'^2}_{1000})^{\frac{1}{2}}$, for the 72-hour forecast is plotted in Fig. 20. Comparing this to Fig. 5a of Blackmon, et al. (1979) which shows the ratio of the 500 to 1000 mb climatological standard deviations, one notes the similarity in the shape of the two fields. This similarity implies a close associa-

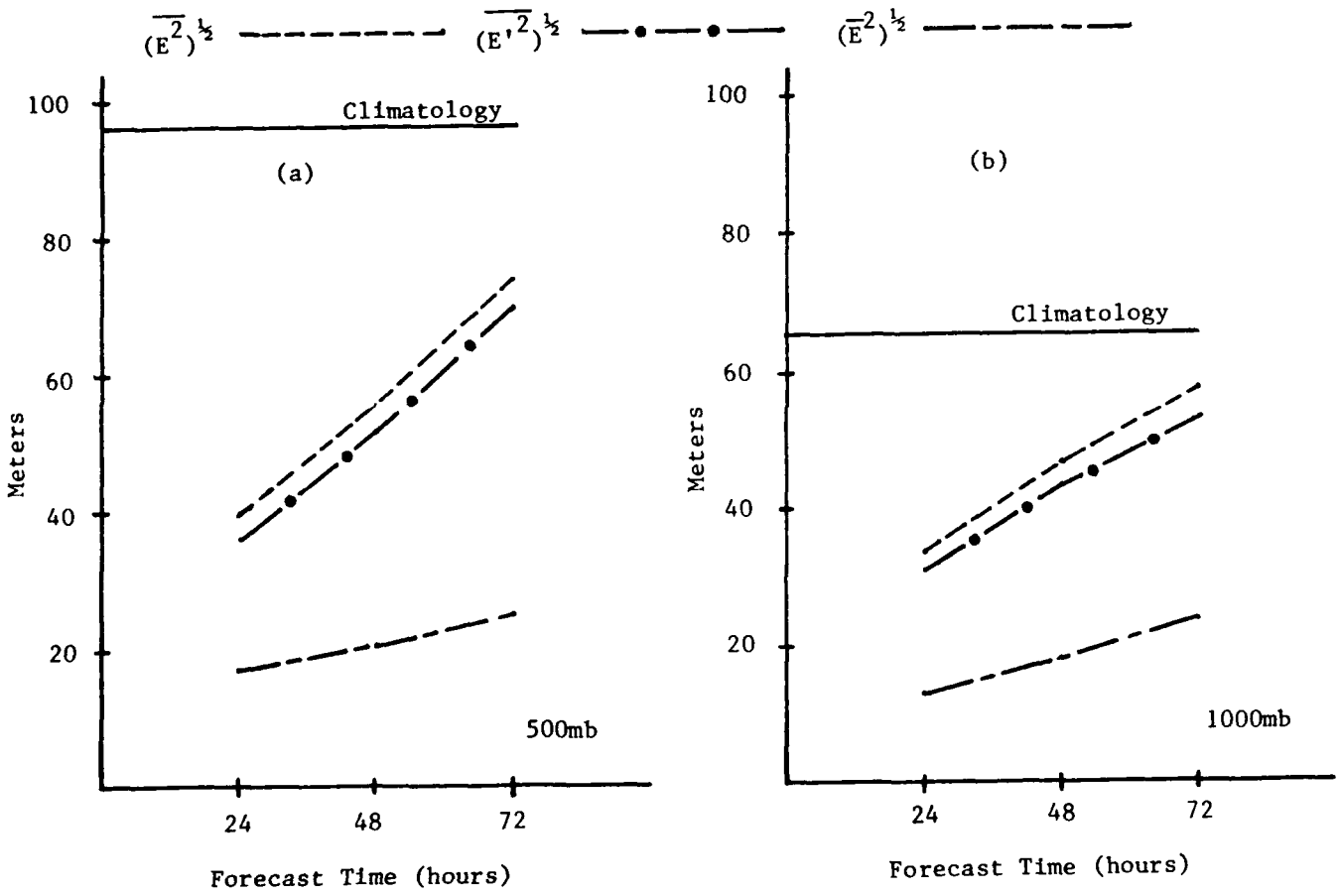


Fig. 18. Area-averaged r.m.s. error, standard deviation of the error, and mean error for (a) 500mb and (b) 1000mb forecasts.

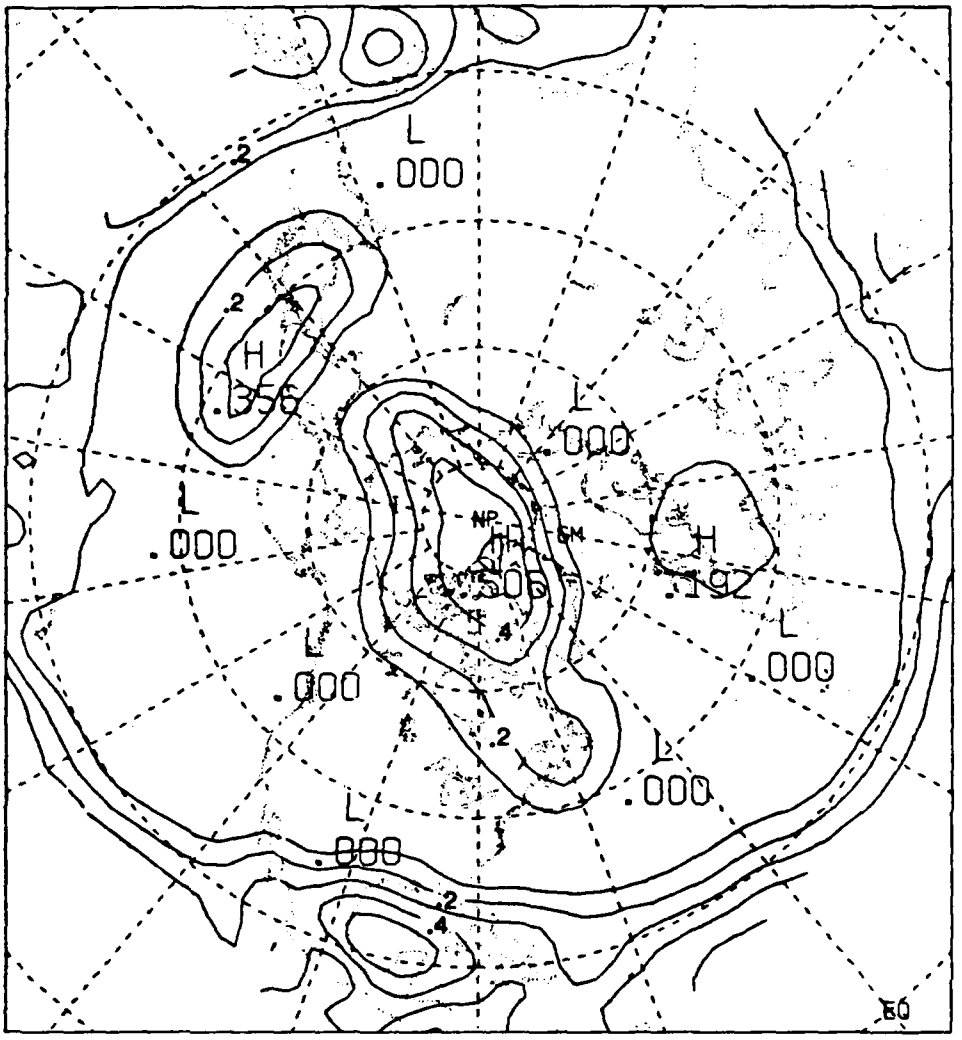


Fig. 19. Percent of total variance of the error due to the bias for the 500mb 72-hour forecast; contour interval 0.1.

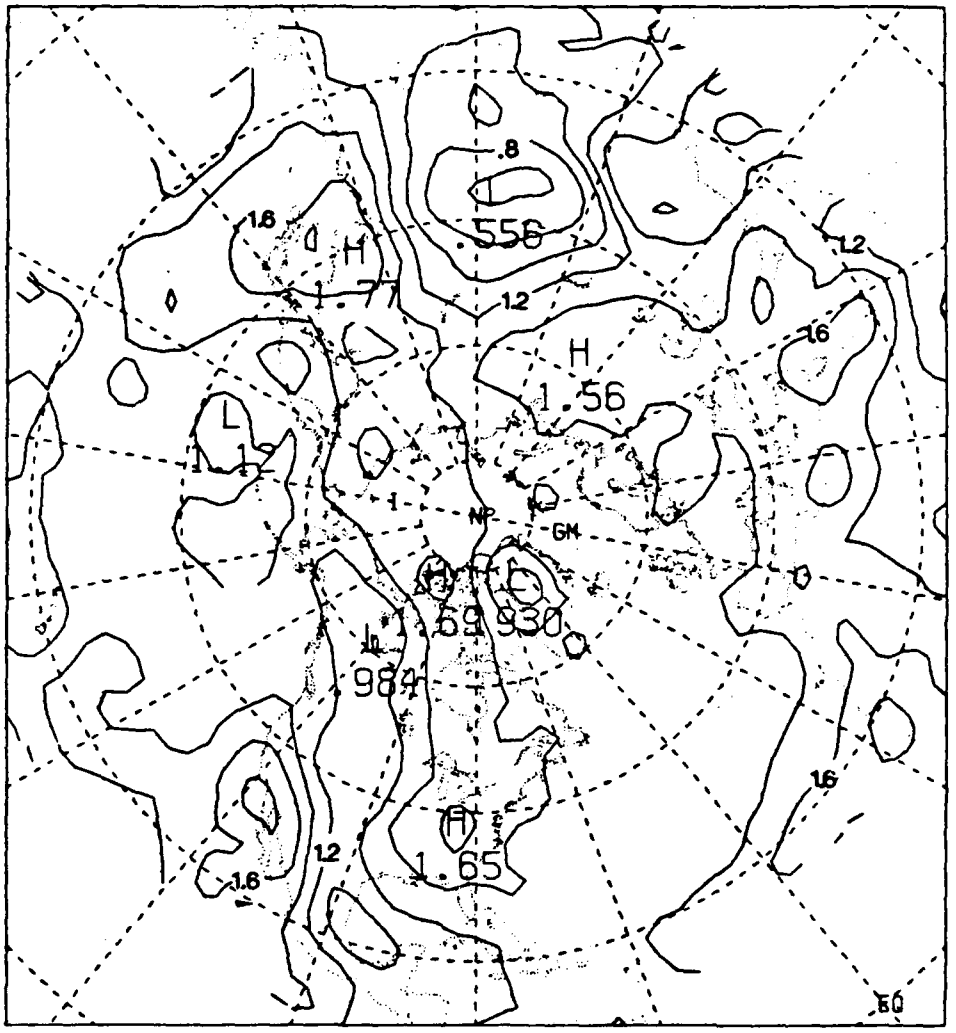


Fig. 20. Ratio of the 500 to 1000mb standard deviation of the error for the 72-hour forecast; contour interval 0.2.

tion between the vertical structure of the error and that of climatology.

The correlation coefficient between the 500 and 1000 mb error for the 24- and 72-hour forecast was computed for each grid point over the six winter seasons. Results are shown in Fig. 21. The 72-hour error correlation (Fig. 21a) shows a remarkable resemblance to the correlation between the 500 and 1000 mb observed height fluctuations [Blackmon, et al. (1979), Fig. 3a] with respect to the shape of the fields with only slight differences in the magnitudes of the maxima and minima centers. The significance of this striking similarity is that the vertical structure of the error field is not random, but it is much like that of real atmospheric fluctuations. The correlation map for the 24-hour forecast error is shown in Fig. 21b. This distribution shows some resemblance to the observed fields but to a much lesser extent. The regional contrasts are not as large as those for the 72-hour forecast errors.

The ratio of the 500 to 1000 mb normalized standard deviation of the error, $(\overline{E'^2}/\overline{O'^2})^{\frac{1}{2}}_{500}/(\overline{E'^2}/\overline{O'^2})^{\frac{1}{2}}_{1000}$, gives a comparison of the model performance at the two levels. It is widely believed that forecasts are more skillful at 500 mb. The area-averaged statistics (Fig. 18) seem to support this idea. Values of the above ratio, averaged over the NMC grid, are .79 for the 24-hour forecasts, .80 for the 48-hour forecasts, and .89 for the 72-hour forecasts. But these ratios are not uniform over the entire grid.

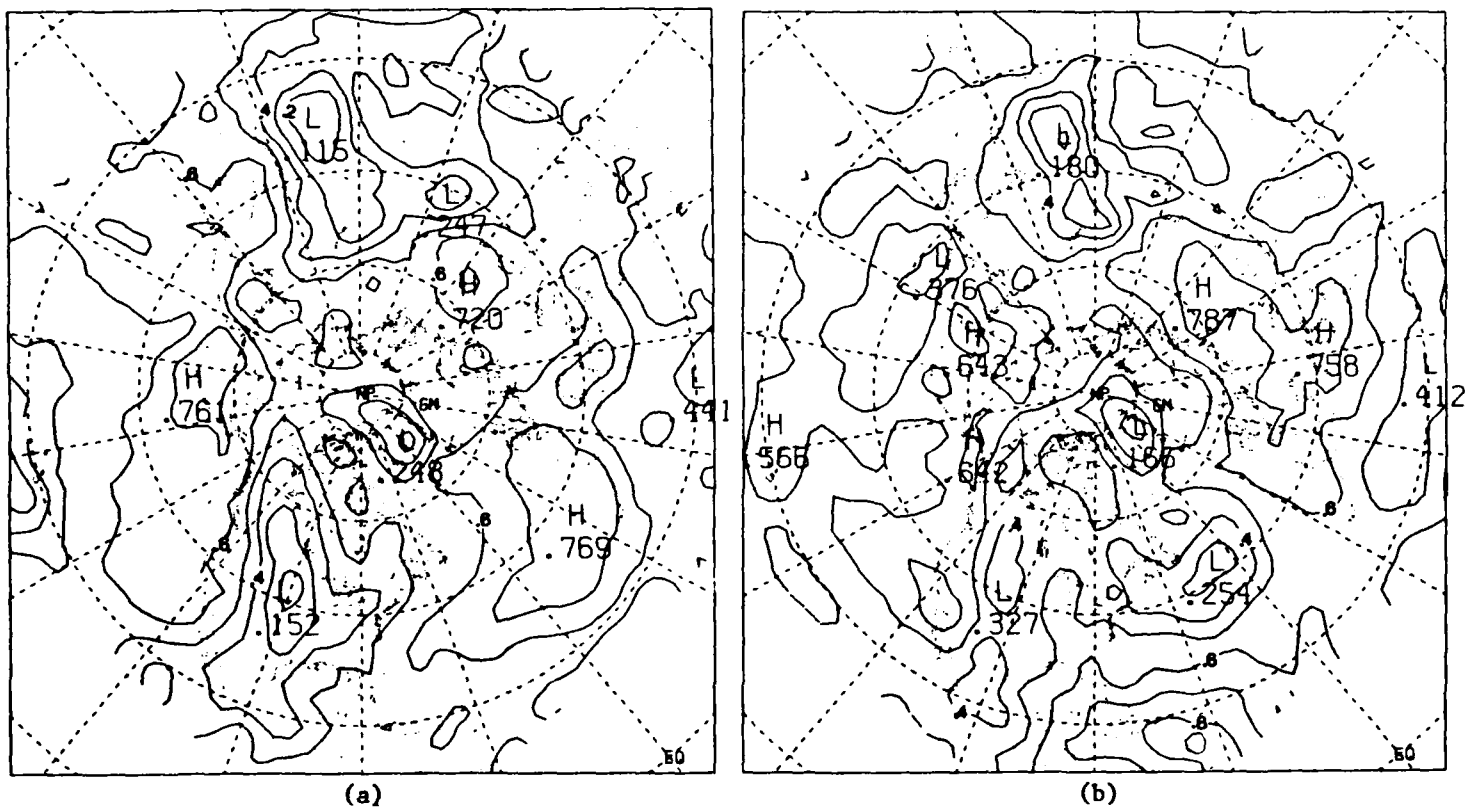


Fig. 21. Correlation coefficients between the 500 and 1000mb mean errors for the (a) 72-hour and (b) 24-hour forecast; contour interval 0.1.

In fact, there are some regions where the model performs better at 1000 mb than at 500 mb. Fig. 22 shows the geographical distribution of this ratio for the 72-hour forecasts. Over the eastern U.S. and eastern Asia, the model performs considerably better at the mid-tropospheric levels than at the ground. However, over the eastern Atlantic and Europe, performance at the two levels tends to be roughly comparable.

d. Systematic Phase Errors

In previous sections it was shown that the PE model is characterized by systematic forecast errors which tend to damp out the mean zonal flow and the forced planetary waves. These forecast errors lead to a progressive weakening of the planetary scale wind field throughout the forecast interval. Taking the gradient of the height error from the center to the southern edge of the positive bias regions off the eastern coasts of Northern American and Asian continents (Fig. 10c), the error in the geostrophic wind was calculated to be 3.5 and 7.5 m/s (or 6.3% and 8.3%, respectively) slower than the observed wind.

It is of interest to determine whether this erroneous weakening of the planetary scale flow with time contributes to the non-systematic component of the error, $(\overline{E'^2})^{\frac{1}{2}}$, by virtue of its effect in the phase speeds of transient weather systems. One would expect the resulting underestimate of the phase speeds to be most serious in the jetstream regions where the systematic underestimate of the zonal wind is larger and the transient variability is also large.

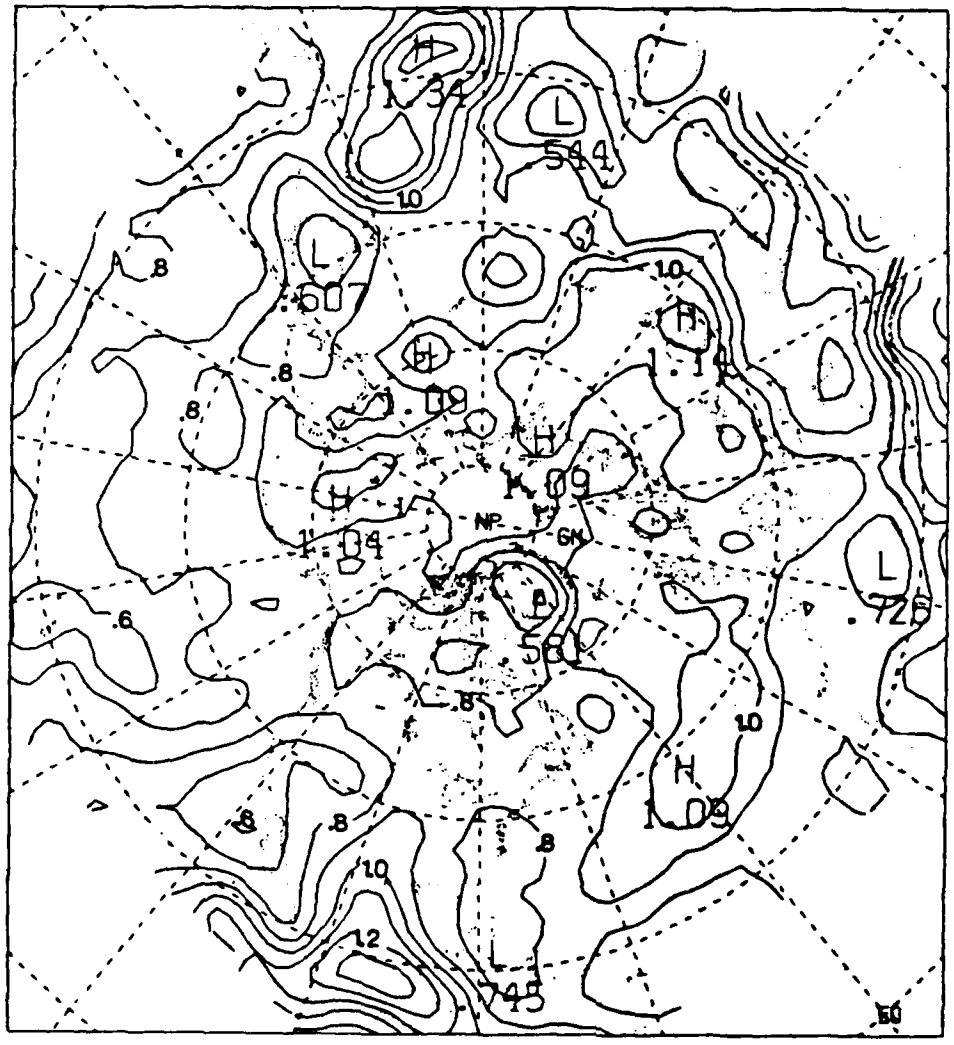


Fig. 22. Ratio of the 500 to 1000mb normalized standard deviation of the error for the 72-hour forecast; contour interval 0.1.

In order to determine the seriousness of underestimated phase speeds as a potential source of error in the forecast, the cross spectra was computed between the 72-hour forecasts and observations at the nine points indicated by the dots in Fig. 10c, which include two jetstream locations. Cross spectra were computed for each winter season individually and then averaged over the six seasons. The time series used were the 500 mb 72-hour forecast and verifying observations (00Z data only). Each season contained ninety data points spaced 24 hours apart. In order to obtain reliable spectral estimates, a lag of 5 days was used giving 18 degrees of freedom. The results of the cross correlation for four of the nine grid points are shown in Table 1. In all spectral bands for all nine grid points including the results of regions not shown, there were negative phase errors. But because the squared coherence is low at the higher frequencies, the confidence in these phase estimates is marginal. Using tables to determine the reliability of the spectral estimates, the squared coherence must be greater than .239(.162) for the 99%(95%) confidence level. So, at most of the grid points, confidence can be given to phase angles through the first three or four phase estimates. In these lowest frequency spectral bands, phase lags are generally between -0.01 to -0.11 cycles. With such small phase lags along with the fact that the forecasted mean zonal wind was only 6 to 8% too slow, it can be concluded that the systematic phase lage contributes only a minor part to $(\overline{E'^2})^{\frac{1}{2}}$.

TABLE 1. Cross spectral analysis between forecast and observation time series at selected grid points.

NORTHERN PACIFIC STANDARD DEVIATION MAXIMUM (170W/50N)

FREQUENCY (CYC/DAY)	FORECAST POWER SPECTRUM	OBSERVATION POWER SPECTRUM	COSPECTRUM	QUADRATURE SPECTRUM	COHERENCE SQUARED	PHASE (CYC)
0	20121.2	25925.7	20095.8	-754.4	.78	-.01
.10	11354.3	15160.7	10819.8	-943.5	.69	-.01
.20	3316.6	3853.4	1965.1	-668.8	.34	-.05
.30	2195.7	1550.8	690.4	-419.9	.19	-.09
.40	1517.1	1004.8	380.2	-148.0	.11	-.06
.50	1234.6	878.7	355.7	6.5	.12	.00

NORTHERN ATLANTIC STANDARD DEVIATION MAXIMUM (31W/53N)

FREQUENCY (CYC/DAY)	FORECAST POWER SPECTRUM	OBSERVATION POWER SPECTRUM	COSPECTRUM	QUADRATURE SPECTRUM	COHERENCE SQUARED	PHASE (CYC)
0	37432.2	42140.7	36472.4	-1867.6	.85	-.01
.10	21424.5	23864.7	19977.1	-2331.9	.79	-.02
.20	5233.5	6206.6	3903.5	-1504.0	.54	-.06
.30	2163.1	2632.7	998.3	- 783.3	.28	-.11
.40	1485.7	1396.5	62.8	- 417.2	.09	-.23
.50	1383.6	1121.2	-248.8	- 148.8	.05	-.41

TABLE 1 (CONTINUED)

ASIAN JET MAXIMUM (149E/29N)

FREQUENCY (CYC/DAY)	FORECAST POWER SPECTRUM	OBSERVATION POWER SPECTRUM	COSPECTRUM	QUADRATURE SPECTRUM	COHERENCE SQUARED	PHASE (CYC)
0	3309.0	5476.6	3217.4	-126.4	.57	-.01
.10	1959.3	3075.4	1709.4	-135.3	.49	-.01
.20	618.6	717.5	304.1	- 43.2	.21	-.02
.30	334.7	262.6	104.4	- 14.6	.13	-.02
.40	235.0	154.7	34.7	- 16.6	.04	-.07
.50	222.9	131.6	6.5	- 5.0	.00	-.11

NORTH AMERICAN JET MAXIMUM (80W/30N)

FREQUENCY (CYC/DAY)	FORECAST POWER SPECTRUM	OBSERVATION POWER SPECTRUM	COSPECTRUM	QUADRATURE SPECTRUM	COHERENCE SQUARED	PHASE (CYC)
0	4929.7	4973.7	4184.7	-258.4	.72	-.01
.10	2761.7	3161.2	2295.1	-295.8	.61	-.02
.20	698.2	1047.3	404.9	-135.8	.32	-.05
.30	315.8	386.3	144.9	- 51.3	.19	-.05
.40	216.9	195.3	47.6	- 29.2	.07	-.09
.50	205.5	155.0	27.2	- 4.3	.02	-.02

7. Discussion

The biases found in this study at the 500 mb level were quite small and insignificant over most of the grid when compared to the r.m.s. error. It is only around the periphery of the grid, over the pole, and in the standing wave troughs that the bias contributes a significant part of the error. Empirical corrections after each forecast would eliminate the bias but would probably not be noticed in the model simulation of atmospheric disturbances. Improvement in the r.m.s. error statistic would be slight. The important aspect of this finding is that the PE model flattens out the climatological mean fields at both the surface and mid-tropospheric levels.

The day-to-day variability of the error fields is closely associated with the day-to-day variability of the observed field. This resemblance, as seen by comparison of the standard deviation of the error to the climatological standard deviation, is present throughout the forecast period. The rate of growth of the non-systematic error and the amplitude of the error fluctuations are largest over the northern oceanic storm track regions while they are smallest over the low latitude tropics. However, when the amplitudes of the non-systematic error fluctuations are compared to the climatological variability, the degradation in the forecast, as manifested by the pole-to-equator increase of the normalized error fields, is greatest approaching the equatorward edge of the grid.

The vertical structure of the fluctuations in the error field closely resemble the vertical structure of real atmospheric fluctuations. The 500 to 1000 mb errors are highly correlated in geographical locations where the observed 500 and 1000 mb surfaces are highly correlated; the errors are poorly correlated where observed 500 and 1000 mb are poorly correlated.

REFERENCES

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. J. Atmos. Sci., 33, 1607-1623.

_____ , R. A. Madden, J. M. Wallace, and D. S. Gutzler, 1979: Geographical variations in the vertical structure of geopotential height fluctuations. J. Atmos. Sci., 36, 2450-2466.

_____ , J. M. Wallace, N. C. Lau, and S. L. Mullen, 1977: An observational study of the Northern Hemisphere wintertime circulation. J. Atmos. Sci., 34, 1040-1053.

Baumheftner, D., P. Downey, 1978: Forecasting intercomparisons from three numerical weather prediction models. Mon. Wea. Rev., 106, 1245-2379.

Conference Proceedings, NCAR Summer Colloquium, 1976: Weather Forecasting and Weather Forecasts: Models, Systems and Users, Summer Colloquium, 1976.

Cressman, G. P., 1959: An operational objective analysis system. Mon. Wea. Rev., 87, 367-374.

_____ , 1963: A three-level model suitable for daily numerical forecasting. Tech. Memo No. 22, National Meteorological Center, Weather Bureau, Washington, D. C., 22 pp.

Fawcett, E. B., 1969: Systematic errors in operational baroclinic prognoses at the National Meteorological Center. Mon. Wea. Rev., 97, 670-682.

_____ , 1977: Current capabilities at the National Weather Service's National Meteorological Center. Bull. Amer. Meteor. Soc., 58, 143-149.

Haltiner, G. J., and R. T. Williams, 1975: Some recent advances in numerical weather prediction. Mon. Wea. Rev., 103, 571-590.

Howcraft, J., 1971: Local forecast model: Present status and preliminary verification. Office Note No. 50, National Meteorological Center, National Weather Service, Washington, D. C., 22 pp.

Leith, C., 1976: Statistical-dynamical forecasting methods. Lecture, NCAR Summer Colloquium, June 16-18, Boulder, Colorado.

McPherson, R., K. Bergman, R. Kistler, G. Rosch and D. Gordon, 1979: The NMC operational global data assimilation system. Mon. Wea. Rev., 107, 1445-1461.

National Weather Service Technical Procedures Bulletin, No. 105, 1974: A Description of the Flattery Global Analysis Method - No. 1. National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, D. C.

_____ , No. 249, 1978: Optimum Interpolation in Final Cycle. National Oceanic and Atmospheric Administration, U.S.

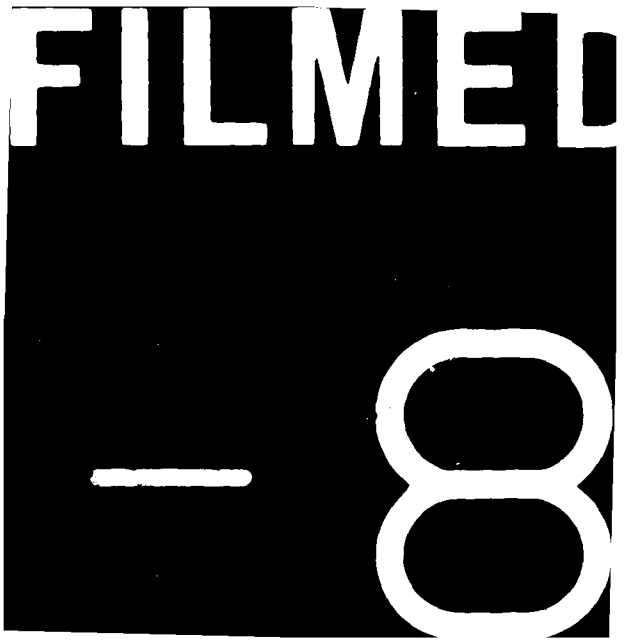
Department of Commerce, Washington, D. C.

Oort, A. H., 1977: Adequacy of the rawinsonde network for global circulation studies tested through numerical model output. Mon. Wea. Rev., 106, 174-195.

Shuman, F. G., 1978: Numerical weather prediction. Bull. Amer. Meteor. Soc., 59, 5-17.

_____ , J. B. Hovermale, 1968: An operational six-layer primative equation model. J. Appl. Meteor., 7, 525-547.

Teweles, S., Jr., and H. B. Wobus, 1954: Verification of prognostic charts. Bull. Amer. Meteor. Soc., 35, 455-463.


FILMED
8